Second Term Improvement to Generalised Linear Mixed Model Asymptotics

By Luca Maestrini[1], Aishwarya Bhaskaran[2] and Matt P. Wand[3]

[1]*The Australian National University,* [2]*Macquarie University and* [3]*University of Technology Sydney*

8th May, 2023

Abstract

A recent article on generalised linear mixed model asymptotics, Jiang *et al.* (2022), derived the rates of convergence for the asymptotic variances of maximum likelihood estimators. If m denotes the number of groups and n is the average within-group sample size then the asymptotic variances have orders m^{-1} and $(mn)^{-1}$, depending on the parameter. We extend this theory to provide explicit forms of the $(mn)^{-1}$ second terms of the asymptotically harder-to-estimate parameters. Improved accuracy of studentised confidence intervals is one consequence of our theory.

Keywords: Longitudinal data analysis, Maximum likelihood estimation, Studentisation.

1 Introduction

Generalised linear mixed models are a vehicle for regression analysis of grouped data with non-Gaussian responses such as counts and categorical labels. Until recently, the precise asymptotic behaviours of the conditional maximum likelihood estimators were not known for these models. Jiang *et al.* (2022) derived leading term asymptotic variances and showed them have orders m^{-1} and $(mn)^{-1}$, depending on the parameter, where m is the number of groups and n is the average within-group sample size. The main contribution of this article is to extend the asymptotic variance and covariance approximations to terms in $(mn)^{-1}$ for *all* parameters. This constitutes *second term improvement* to generalized linear mixed model asymptotics. The potential statistical payoffs are improved accuracy of confidential intervals, hypothesis tests, sample size calculations and optimal design.

The essence of generalized linear mixed models is the extension of general linear models via the addition of random effects that allow for the handling of correlations arising from repeated measures. There are numerous types of random effect structures. The most common is the two-level nested structure, corresponding to repeated measures within each of m distinct groups. This version of generalised linear mixed models, with frequentist inference via maximum likelihood and its quasi-likelihood extension, is our focus here. Overviews of generalised linear mixed models are provided by books such as Jiang & Nguyen (2021), McCulloch *et al.* (2008) and Stroup (2013).

Suppose that a fixed effects parameter in a two-level generalised linear mixed model is accompanied by a random effect. Jiang *et al.* (2022) showed that the variance of its maximum likelihood estimator, conditional on the predictor data, is asymptotic to $C_1 m^{-1}$ for some deterministic constant C_1 that depends on the true model parameter values. The crux of this article is to extend the asymptotic variance approximation to $C_1 m^{-1} + C_2(mn)^{-1}$ for an additional deterministic constant C_2. We derive the explicit form of C_2 for two-level nested generalised linear mixed models for both maximum likelihood and maximum quasi-likelihood situations. Even though, in general, C_2 does not have a succinct form it is still usable in that operations such as studentisation are straightforward and result in improvements in statistical utility.

For two-level nested mixed models, $(mn)^{-1}$ is the best possible rate of convergence for the asymptotic variance of the estimator of a model parameter. Such a rate is achieved by maximum likelihood estimators of fixed effects parameters unaccompanied by random effects and dispersion parameters (e.g. Bhaskaran & Wand, 2023). The current article closes the problem of obtaining the precise asymptotic forms of the variances, up to terms in $(mn)^{-1}$, for estimation of *all* model parameters.

Section 2 describes the model under consideration and corresponding maximum estimators. Our second term improvement results are presented in Section 3. Section 4 describes statistical utility due to the new asymptotic results. We present some corroborating numerical results in Section 5. A supplement to this article contains derivational details.

2 Model Description and Maximum Likelihood Estimation

Consider the class of two-parameter exponential family of density, or probability mass, functions with generic form

$$p(y;\eta,\phi)=\exp[\{y\eta-b(\eta)+c(y)\}/\phi+d(y,\phi)]h(y) \tag{1}$$

where η is the *natural parameter* and $\phi>0$ is the *dispersion parameter*. Examples include the Gaussian density for which $b(x)=\frac{1}{2}x^2$, $c(x)=-\frac{1}{2}x^2$, $d(x_1,x_2)=-\frac{1}{2}\log(2\pi x_2)$ and $h(x)=I(x\in\mathbb{R})$ and the Gamma density function for which $b(x)=-\log(-x)$, $c(x)=\log(x)$, $d(x_1,x_2)=-\log(x_1)-\log(x_2)/x_2-\log\Gamma(1/x_2)$ and $h(x)=I(x>0)$. Here $I(\mathcal{P})=1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P})=0$ if $\mathcal{P}$ is false. The Binomial and Poisson probability mass functions are also special cases of (1) but with ϕ fixed at 1. When (1) is used in regression contexts a common modelling extension for count and proportion responses, usually to account for overdispersion, is to remove the $\phi=1$ restriction and replace it with $\phi>0$. In these circumstances

$$\{y\eta-b(\eta)+c(y)\}/\phi+d(y,\phi) \tag{2}$$

is labelled a *quasi-likelihood function* since it is not the logarithm of a probability mass function for $\phi\neq 1$. We use the more general quasi-likelihood terminology for the remainder of this article.

Consider, for observations of the random pairs $(\boldsymbol{X}_{ij},Y_{ij})$, $1\le i\le m$, $1\le j\le n_i$, generalised linear mixed models of the form,

$$\begin{array}{l}Y_{ij}|\boldsymbol{X}_{ij},\boldsymbol{U}_i \text{ independent having quasi-likelihood function (2) with natural parameter}\\ \left(\boldsymbol{\beta}^0+\begin{bmatrix}\boldsymbol{U}_i\\ \mathbf{0}\end{bmatrix}\right)^T\boldsymbol{X}_{ij} \text{ such that the } \boldsymbol{U}_i \text{ are independent } N(\mathbf{0},\boldsymbol{\Sigma}^0) \text{ random vectors.}\end{array} \tag{3}$$

The $\boldsymbol{X}_{ij}$ are $d_{\text{F}}\times 1$ random vectors corresponding to predictors. The $\boldsymbol{U}_i$ are $d_{\text{R}}\times 1$ unobserved random effects vectors, where $d_{\text{R}}\le d_{\text{F}}$. Under this set-up the first d_{R} entries of the $\boldsymbol{X}_{ij}$ are partnered by a random effect. The remaining entries correspond to predictors that have a fixed effect only. We assume that the $\boldsymbol{X}_{ij}$ and $\boldsymbol{U}_i$, for $1\le i\le m$ and $1\le j\le n_i$, are totally independent, with the $\boldsymbol{X}_{ij}$ each having the same distribution as the $d_{\text{F}}\times 1$ random vector $\boldsymbol{X}$ and the $\boldsymbol{U}_i$ each having the same distribution as the $d_{\text{R}}\times 1$ random vector $\boldsymbol{U}$.

For any $\boldsymbol{\beta}$ ($d_{\text{F}}\times 1$) and $\boldsymbol{\Sigma}$ ($d_{\text{R}}\times d_{\text{R}}$) that is symmetric and positive definite and conditional on the $\boldsymbol{X}_{ij}$ data, the quasi-likelihood is

$$\begin{aligned}\ell(\boldsymbol{\beta},\boldsymbol{\Sigma})=&\sum_{i=1}^m\sum_{j=1}^{n_i}[\{Y_{ij}(\boldsymbol{\beta}^T\boldsymbol{X}_{ij}+c(Y_{ij})\}/\phi+d(Y_{ij},\phi)]-\frac{m}{2}\log|2\pi\boldsymbol{\Sigma}|\\ &+\sum_{i=1}^m\log\int_{\mathbb{R}^{d_{\text{R}}}}\exp\left[\frac{1}{\phi}\sum_{j=1}^{n_i}\left\{Y_{ij}\begin{bmatrix}\boldsymbol{u}\\ \mathbf{0}\end{bmatrix}^T\boldsymbol{X}_{ij}-b\left(\left(\boldsymbol{\beta}+\begin{bmatrix}\boldsymbol{u}\\ \mathbf{0}\end{bmatrix}\right)^T\boldsymbol{X}_{ij}\right)\right\}-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right]d\boldsymbol{u}.\end{aligned}$$

The maximum quasi-likelihood estimator of $(\boldsymbol{\beta}^0,\boldsymbol{\Sigma}^0)$ is

$$(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Sigma}})=\underset{\boldsymbol{\beta},\boldsymbol{\Sigma}}{\text{argmax}}\,\ell(\boldsymbol{\beta},\boldsymbol{\Sigma}).$$

Suppose that $d_{\text{F}}>d_{\text{R}}$ and consider the partition $\boldsymbol{\beta}=[\boldsymbol{\beta}_{\text{A}}^T\ \boldsymbol{\beta}_{\text{B}}^T]^T$ of the fixed effects parameter vector, where $\boldsymbol{\beta}_{\text{A}}$ is $d_{\text{R}}\times 1$ and $\boldsymbol{\beta}_{\text{B}}$ is $(d_{\text{F}}-d_{\text{R}})\times 1$. The $d_{\text{F}}=d_{\text{R}}$ boundary case is such that $\boldsymbol{\beta}_{\text{B}}$ is null. Also, let

$\boldsymbol{\mathcal{X}} \equiv \{\boldsymbol{X}_{ij} : 1 \le i \le m,\ 1 \le j \le n_i\}$. Theorem 1 of Jiang *et al.* (2022) implies that, under some mild conditions, the covariance matrices of $\widehat{\boldsymbol{\beta}}_{\rm A}$, $\widehat{\boldsymbol{\beta}}_{\rm B}$ and vech$(\widehat{\boldsymbol{\Sigma}})$ have leading term behaviour given by

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}_{\rm A}|\boldsymbol{\mathcal{X}}\big) = \frac{\boldsymbol{\Sigma}^0\{1+o_p(1)\}}{m},\ \ \text{Cov}\big(\widehat{\boldsymbol{\beta}}_{\rm B}|\boldsymbol{\mathcal{X}}\big) = \frac{\phi\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\rm B}}\{1+o_p(1)\}}{mn},\ \ \text{where}\ \ n \equiv \frac{1}{m}\sum_{i=1}^m n_i, \tag{4}$$

and

$$\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{2\boldsymbol{D}_{d_{\rm R}}^{+}(\boldsymbol{\Sigma}^0\otimes\boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\rm R}}^{+T}\{1+o_p(1)\}}{m}. \tag{5}$$

Here $\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\rm B}}$ is a $(d_{\rm F}-d_{\rm R})\times(d_{\rm F}-d_{\rm R})$ matrix that depends on $\boldsymbol{\beta}$ and the $(\boldsymbol{X},\boldsymbol{U})$ distribution, $\boldsymbol{D}_{d_{\rm R}}$ is the matrix of zeroes and ones such that $\boldsymbol{D}_{d_{\rm R}}\text{vech}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A})$ for all $d_{\rm R}\times d_{\rm R}$ symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{D}_{d_{\rm R}}^{+} = (\boldsymbol{D}_{d_{\rm R}}^{T}\boldsymbol{D}_{d_{\rm R}})^{-1}\boldsymbol{D}_{d_{\rm R}}^{T}$ is the Moore-Penrose inverse of $\boldsymbol{D}_{d_{\rm R}}$. The theory of Jiang *et al.* (2022) also indicates a degree of asymptotic orthogonality between $\boldsymbol{\beta}_{\rm A}$ and $\boldsymbol{\beta}_{\rm B}$ in that $E\big\{(\widehat{\boldsymbol{\beta}}_{\rm A}-\boldsymbol{\beta}_{\rm A}^0)(\widehat{\boldsymbol{\beta}}_{\rm B}-\boldsymbol{\beta}_{\rm B}^0)^T|\boldsymbol{\mathcal{X}}\big\}$ has $O_p\{(mn)^{-1}\}$ entries, which implies that the correlations between the entries of $\widehat{\boldsymbol{\beta}}_{\rm A}$ and $\widehat{\boldsymbol{\beta}}_{\rm B}$ are asymptotically negligible.

The leading term approximations of the variability in $\widehat{\boldsymbol{\beta}}_{\rm A}$ and vech$(\widehat{\boldsymbol{\Sigma}})$, given by (4) and (5), are somewhat crude. Unlike the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_{\rm B}$, they do not show the effect of the average within-group sample size n. In the next section we investigate their second term improvements.

3 Two-Term Asymptotic Covariance Results

We define the two-term asymptotic covariance matrix problem to be the determination of the *unique deterministic* matrices $\boldsymbol{M}_{\boldsymbol{\beta}}$ and $\boldsymbol{M}_{\boldsymbol{\Sigma}}$ such that

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\begin{bmatrix}\boldsymbol{\Sigma}^0 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix} + \frac{\boldsymbol{M}_{\boldsymbol{\beta}}\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{2\boldsymbol{D}_{d_{\rm R}}^{+}(\boldsymbol{\Sigma}^0\otimes\boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\rm R}}^{+T}}{m} + \frac{\boldsymbol{M}_{\boldsymbol{\Sigma}}\{1+o_p(1)\}}{mn}$$

under reasonably mild conditions.

An example for which a solution to the two-term asymptotic covariance problem can be expressed relatively simply is the $d_{\rm F}=2$, $d_{\rm R}=1$ Poisson quasi-likelihood special case of (3), with parameters

$$\boldsymbol{\beta} = (\beta_0,\beta_1) \quad \text{and} \quad \boldsymbol{\Sigma} = \sigma^2 \quad \text{and predictor variable} \quad \boldsymbol{X} = \begin{bmatrix}1\\ X\end{bmatrix}$$

for a scalar random variable X. Define

$$a_1\big(\beta_0,\beta_1,\sigma^2\big) \equiv e^{\beta_0+\sigma^2/2}\big[E(X^2e^{\beta_1X})E(e^{\beta_1X}) - \{E(Xe^{\beta_1X})\}^2\big]$$

and

$$a_2(\beta_1,\sigma^2) \equiv \frac{e^{\sigma^2}E\big(X^2e^{\beta_1X}\big)E\big(e^{\beta_1X}\big) + \big(1-e^{\sigma^2}\big)E\big\{\big(Xe^{\beta_1X}\big)\big\}^2}{E\big(e^{\beta_1X}\big)}.$$

Then the two-term covariance matrix of $(\widehat{\beta}_0,\widehat{\beta}_1)$ is

$$\text{Cov}\left(\left[\begin{array}{c}\widehat{\beta}_0\\ \widehat{\beta}_1\end{array}\right]\middle|\boldsymbol{\mathcal{X}}\right) = \frac{1}{m}\left[\begin{array}{cc}(\sigma^2)^0 & 0\\ 0 & 0\end{array}\right] + \frac{\phi\{1+o_p(1)\}}{a_1\big(\beta_0^0,\beta_1^0,(\sigma^2)^0\big)\,mn}\left[\begin{array}{cc}a_2\big(\beta_1^0,(\sigma^2)^0\big) & -E\big(Xe^{\beta_1^0X}\big)\\ -E\big(Xe^{\beta_1^0X}\big) & E\big(e^{\beta_1^0X}\big)\end{array}\right].$$

In other words, for this simple example, the solution for $\boldsymbol{M}_{\boldsymbol{\beta}}$ is

$$\boldsymbol{M}_{\boldsymbol{\beta}} = \frac{\phi}{a_1\big(\beta_0^0,\beta_1^0,(\sigma^2)^0\big)}\left[\begin{array}{cc}a_2\big(\beta_1^0,(\sigma^2)^0\big) & -E\big(Xe^{\beta_1^0X}\big)\\ -E\big(Xe^{\beta_1^0X}\big) & E\big(e^{\beta_1^0X}\big)\end{array}\right].$$

Studentisation of the two-term asymptotic covariance matrix for obtaining confidence intervals and Wald hypothesis tests is straightforward. For example, $E(X^2 e^{\beta_1^0 X})$ can be replaced by the estimator

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}^2 e^{\widehat{\beta}_1 X_{ij}}.$$

This practical aspect is discussed in depth in Section 4.

The remainder of this section is concerned with the *theoretical* problem of obtaining the forms of $\boldsymbol{M}_{\boldsymbol{\beta}}$ and $\boldsymbol{M}_{\boldsymbol{\Sigma}}$ for model (3) in general. The achievement of this goal has turned out to be quite challenging. The score asymptotic approximation approach used in Jiang *et al.* (2022) requires higher numbers of terms to obtain valid two-term covariance matrix approximations. Some of these terms can only be expressed using three-dimensional arrays rather than with matrices. Succinct statement of $\boldsymbol{M}_{\boldsymbol{\beta}}$ and $\boldsymbol{M}_{\boldsymbol{\Sigma}}$ is only possible with well-designed nested function notation. A novel notation for multiplicative combining of three-dimensional arrays with compatible matrices is also beneficial. The next subsection focusses on these notational aspects.

3.1 Notation for the Main Result

Let $\boldsymbol{\mathcal{A}}$ be a $d_1 \times d_2 \times d_3$ array and $\boldsymbol{M}$ be a $d_1 \times d_2$ matrix. Then we let

$$\boldsymbol{\mathcal{A}}\bigstar\boldsymbol{M} \quad \text{denote the } d_3 \times 1 \text{ vector with } t\text{th entry given by } \quad \sum_{r=1}^{d_1}\sum_{s=1}^{d_2}(\boldsymbol{\mathcal{A}})_{rst}(\boldsymbol{M})_{rs}. \tag{6}$$

Next, for $\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^0)$, define

$$\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U}) \equiv E\Big\{b''\big((\boldsymbol{\beta}_{\text{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+(\boldsymbol{\beta}_{\text{B}}^0)^T\boldsymbol{X}_{\text{B}}\big)\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T|\boldsymbol{U}\Big\},$$

$$\boldsymbol{\Omega}_{\text{AB}}(\boldsymbol{U}) \equiv E\Big\{b''\big((\boldsymbol{\beta}_{\text{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+(\boldsymbol{\beta}_{\text{B}}^0)^T\boldsymbol{X}_{\text{B}}\big)\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T|\boldsymbol{U}\Big\}$$

and

$$\boldsymbol{\Omega}_{\text{BB}}(\boldsymbol{U}) \equiv E\Big\{b''\big((\boldsymbol{\beta}_{\text{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+(\boldsymbol{\beta}_{\text{B}}^0)^T\boldsymbol{X}_{\text{B}}\big)\boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{B}}^T|\boldsymbol{U}\Big\}.$$

Also let $\boldsymbol{\Omega}'_{\text{AAA}}(\boldsymbol{U})$ be the $d_{\text{R}} \times d_{\text{R}} \times d_{\text{R}}$ array with (r,s,t) entry equal to

$$E\Big\{b'''\big((\boldsymbol{\beta}_{\text{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+(\boldsymbol{\beta}_{\text{B}}^0)^T\boldsymbol{X}_{\text{B}}\big)(\boldsymbol{X}_{\text{A}})_r(\boldsymbol{X}_{\text{A}})_s(\boldsymbol{X}_{\text{A}})_t\big|\boldsymbol{U}\Big\}.$$

and $\boldsymbol{\Omega}'_{\text{AAB}}(\boldsymbol{U})$ be the $d_{\text{R}} \times d_{\text{R}} \times (d_{\text{F}}-d_{\text{R}})$ array with (r,s,t) entry equal to

$$E\Big\{b'''\big((\boldsymbol{\beta}_{\text{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\text{A}}+(\boldsymbol{\beta}_{\text{B}}^0)^T\boldsymbol{X}_{\text{B}}\big)(\boldsymbol{X}_{\text{A}})_r(\boldsymbol{X}_{\text{A}})_s(\boldsymbol{X}_{\text{B}})_t\big|\boldsymbol{U}\Big\}.$$

Define the random vectors:

$$\psi_1(\boldsymbol{U}) \equiv \text{vech}(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T), \quad \psi_2(\boldsymbol{U}) \equiv \boldsymbol{\Omega}'_{\text{AAA}}(\boldsymbol{U})\bigstar\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1},$$

$$\psi_3(\boldsymbol{U}) \equiv \boldsymbol{\Omega}'_{\text{AAB}}(\boldsymbol{U})\bigstar\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1} \text{ and } \psi_4(\boldsymbol{U}) \equiv \boldsymbol{D}_{d_{\text{R}}}^{+}\text{vec}\Big(\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}\boldsymbol{\Sigma}^{-1}\big\{\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T-\boldsymbol{\Sigma}\psi_2(\boldsymbol{U})\boldsymbol{U}^T\big\}\Big).$$

Then define the random matrices:

$$\boldsymbol{\Psi}_5(\boldsymbol{U}) \equiv \boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}\boldsymbol{\Omega}_{\text{AB}}(\boldsymbol{U}), \quad \boldsymbol{\Psi}_6(\boldsymbol{U}) \equiv \boldsymbol{\Omega}_{\text{BB}}(\boldsymbol{U})-\boldsymbol{\Psi}_5(\boldsymbol{U})^T\boldsymbol{\Omega}_{\text{AB}}(\boldsymbol{U}),$$

$$\boldsymbol{\Psi}_7(\boldsymbol{U}) \equiv \boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}, \quad \boldsymbol{\Psi}_8(\boldsymbol{U}) \equiv \boldsymbol{D}_{d_{\text{R}}}^{+}\big[(\boldsymbol{U}\boldsymbol{U}^T)\otimes\{\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}\}\big]\boldsymbol{D}_{d_{\text{R}}}^{+T},$$

$$\text{and} \quad \boldsymbol{\Psi}_9(\boldsymbol{U}) \equiv \psi_1(\boldsymbol{U})\psi_4(\boldsymbol{U})^T+\psi_4(\boldsymbol{U})\psi_1(\boldsymbol{U})^T.$$

Lastly, define the expectation matrices:

$$\boldsymbol{\Lambda}_{\text{AA}} \equiv E\Big\{\boldsymbol{\Psi}_7(\boldsymbol{U})+\boldsymbol{\Psi}_7(\boldsymbol{U})^T-\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}+\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}\psi_2(\boldsymbol{U})\boldsymbol{U}^T+\boldsymbol{U}\psi_2(\boldsymbol{U})^T\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U})^{-1}\Big\},$$

$$\boldsymbol{\Lambda}_{\text{AB}} \equiv E\Big\{\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_5(\boldsymbol{U})+\boldsymbol{U}\psi_2(\boldsymbol{U})^T\boldsymbol{\Psi}_5(\boldsymbol{U})-\boldsymbol{U}\psi_3(\boldsymbol{U})^T\Big\} \quad \text{and}$$

$$\boldsymbol{\Phi} \equiv E\Big(\Big[\boldsymbol{\Psi}_5(\boldsymbol{U})^T\big\{\boldsymbol{\Sigma}^{-1}\boldsymbol{U}+\psi_2(\boldsymbol{U})\big\}-\psi_3(\boldsymbol{U})\Big]\psi_1(\boldsymbol{U})^T\Big).$$

3.2 Assumptions for the Main Result

The main result depends on the following sample size asymptotic assumptions:

The number of groups m diverges to ∞.

The within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \le i \le m$.

The ratio n/m converges to zero.

The last of these conditions is in keeping with the number of groups being large compared with the within-group sample sizes, as often arises in practice. For our asymptotics it ensures that, for the harder-to-estimate parameters, the asymptotic variances of the maximum likelihood estimators have leading terms of the form $C_1 m^{-1} + C_2(mn)^{-1}$. In addition, it ensures that the Fisher information is sufficiently dominant for obtaining asymptotic variances.

We also assume that the $(\boldsymbol{X}, \boldsymbol{U})$ joint distribution is such that all required convergence in probability limits that appear in the deterministic order $(mn)^{-1}$ terms are justified. An example of such a convergence in probability statement is

$$\begin{aligned}
&\frac{1}{n}E\Big(\mathcal{H}_{\mathrm{AB}i}^T \mathcal{H}_{\mathrm{AA}i}^{-1} \mathcal{H}_{\mathrm{AB}i}\Big| \boldsymbol{X}_i\Big) \xrightarrow{p} \boldsymbol{\Omega}_{\mathrm{AB}}(\boldsymbol{U})^T \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1} \boldsymbol{\Omega}_{\mathrm{AB}}(\boldsymbol{U}) \\
&\text{where } \mathcal{H}_{\mathrm{AA}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\big) \boldsymbol{X}_{\mathrm{A}ij} \boldsymbol{X}_{\mathrm{A}ij}^T \\
&\text{and } \mathcal{H}_{\mathrm{AB}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i)^T \boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T \boldsymbol{X}_{\mathrm{B}ij}\big) \boldsymbol{X}_{\mathrm{A}ij} \boldsymbol{X}_{\mathrm{B}ij}^T.
\end{aligned} \tag{7}$$

Assumption (A3) of Jiang *et al.* (2022) provides a moment-type condition that is sufficient for (7) to hold. Also, we assume that the tail behaviour of the $(\boldsymbol{X}, \boldsymbol{U})$ distribution is such that statements concerning the $o_p\{(mn)^{-1}\}$ remainder terms are valid. The determination of sufficient conditions on the $(\boldsymbol{X}, \boldsymbol{U})$ distribution that guarantee the validity of the main result is a tall order, and beyond the scope of this article.

3.3 Statement of the Main Result

Using the notation presented in Section 3.1 and under the assumptions described in Section 3.2, and assuming $d_{\mathrm{F}} > d_{\mathrm{R}}$ we have

$$\begin{aligned}
&\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\begin{bmatrix} \boldsymbol{\Sigma}^0 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} + \frac{\phi}{mn}\begin{bmatrix} \boldsymbol{\Lambda}_{\mathrm{AA}}^{-1} & \boldsymbol{\Lambda}_{\mathrm{AA}}^{-1}\boldsymbol{\Lambda}_{\mathrm{AB}} \\ \boldsymbol{\Lambda}_{\mathrm{AB}}^T\boldsymbol{\Lambda}_{\mathrm{AA}}^{-1} & \boldsymbol{\Lambda}_{\mathrm{AB}}^T\boldsymbol{\Lambda}_{\mathrm{AA}}^{-1}\boldsymbol{\Lambda}_{\mathrm{AB}} + E\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\} \end{bmatrix}^{-1} \{1 + o_p(1)\} \quad \text{and} \\
&\mathrm{Cov}\big(\mathrm{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{2\boldsymbol{D}_{d_{\mathrm{R}}}^{+}(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\mathrm{R}}}^{+T}}{m} \\
&\qquad + \frac{\phi}{mn}\Big(2E\big\{\boldsymbol{\Psi}_9(\boldsymbol{U}) - 2\boldsymbol{\Psi}_8(\boldsymbol{U})\big\} + \boldsymbol{\Phi}^T\big[E\{\boldsymbol{\Psi}_6(\boldsymbol{U})\}\big]^{-1}\boldsymbol{\Phi}\Big)\{1 + o_p(1)\}.
\end{aligned} \tag{8}$$

For the $d_{\mathrm{F}} = d_{\mathrm{R}}$ boundary case the first term of $\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big)$ is simply $\frac{1}{m}\boldsymbol{\Sigma}^0$.

A supplement to this article contains a full derivation of (8).

3.3.1 The Gaussian Response Special Case

In the Gaussian response special case we have $b''(x) = 1$ and $b'''(x) = 0$ and the main result reduces to the following succinct form:

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\begin{bmatrix}\boldsymbol{\Sigma}^0 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix} + \frac{\phi\big\{E(\boldsymbol{X}\boldsymbol{X}^T)\big\}^{-1}\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\begin{aligned}\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) &= \frac{2\boldsymbol{D}^{+}_{d_{\text{R}}}(\boldsymbol{\Sigma}^0\otimes\boldsymbol{\Sigma}^0)\boldsymbol{D}^{+T}_{d_{\text{R}}}}{m}\\ &\quad + \frac{4\phi\boldsymbol{D}^{+}_{d_{\text{R}}}\big[\boldsymbol{\Sigma}^0\otimes\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\big]\boldsymbol{D}^{+T}_{d_{\text{R}}}\{1+o_p(1)\}}{mn}.\end{aligned} \tag{9}$$

We are not aware of any previous appearances of (9) in the wider linear mixed model literature.

4 Utility of the Second Term Improvements

We now describe the utility of (8) in statistical contexts such as inference and design. Improved confidence intervals is a particularly straightforward application, which we treat next.

4.1 Confidence Intervals

For any $\boldsymbol{u}\in\mathbb{R}^{d_{\text{R}}}$, define

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}_{\text{AA}}(\boldsymbol{u}) &\equiv \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^{n_i} b''\Big((\widehat{\boldsymbol{\beta}}_{\text{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij}+\widehat{\boldsymbol{\beta}}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\Big)\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{A}ij}^T,\\ \widehat{\boldsymbol{\Omega}}_{\text{AB}}(\boldsymbol{u}) &\equiv \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^{n_i} b''\Big((\widehat{\boldsymbol{\beta}}_{\text{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij}+\widehat{\boldsymbol{\beta}}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\Big)\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{B}ij}^T\\ \text{and}\quad \widehat{\boldsymbol{\Omega}}_{\text{BB}}(\boldsymbol{u}) &\equiv \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^{n_i} b''\Big((\widehat{\boldsymbol{\beta}}_{\text{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij}+\widehat{\boldsymbol{\beta}}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\Big)\boldsymbol{X}_{\text{B}ij}\boldsymbol{X}_{\text{B}ij}^T.\end{aligned} \tag{10}$$

Then the natural studentisation of $E\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\}$ is

$$\begin{aligned}\widehat{E}\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\} &\equiv E\Big\{\widehat{\boldsymbol{\Omega}}_{\text{BB}}(\boldsymbol{U})-\widehat{\boldsymbol{\Omega}}_{\text{AB}}(\boldsymbol{U})^T\widehat{\boldsymbol{\Omega}}_{\text{AA}}(\boldsymbol{U})^{-1}\widehat{\boldsymbol{\Omega}}_{\text{AB}}(\boldsymbol{U})\big|\boldsymbol{\mathcal{X}}\Big\}\\ &= |2\pi\widehat{\boldsymbol{\Sigma}}|^{-1/2}\int_{d_{\text{R}}}\Big\{\widehat{\boldsymbol{\Omega}}_{\text{BB}}(\boldsymbol{u})-\widehat{\boldsymbol{\Omega}}_{\text{AB}}(\boldsymbol{u})^T\widehat{\boldsymbol{\Omega}}_{\text{AA}}(\boldsymbol{u})^{-1}\widehat{\boldsymbol{\Omega}}_{\text{AB}}(\boldsymbol{u})\Big\}\exp\Big(-\tfrac{1}{2}\boldsymbol{u}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{u}\Big)\,d\boldsymbol{u}.\end{aligned} \tag{11}$$

In the last expression of (11) integration is applied element-wise to each entry of the matrix inside the integral. The natural studentisations of

$$\boldsymbol{\Lambda}_{\text{AA}},\quad \boldsymbol{\Lambda}_{\text{AB}},\quad \boldsymbol{\Phi},\quad E\big\{\boldsymbol{\Psi}_8(\boldsymbol{U})\big\}\quad\text{and}\quad E\big\{\boldsymbol{\Psi}_9(\boldsymbol{U})\big\} \tag{12}$$

are analogous to that for $E\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\}$. The studentisations for the quantities in (12) depend on the functions defined by (10) as well as similar sample counterparts of $\boldsymbol{\Omega}'_{\text{AAA}}(\boldsymbol{U})$ and $\boldsymbol{\Omega}'_{\text{AAB}}(\boldsymbol{U})$. Next define

$$\widehat{\text{Asy.Cov}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{m}\begin{bmatrix}\widehat{\boldsymbol{\Sigma}} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix} + \frac{\widehat{\phi}}{mn}\begin{bmatrix}\widehat{\boldsymbol{\Lambda}}_{\text{AA}}^{-1} & \widehat{\boldsymbol{\Lambda}}_{\text{AA}}^{-1}\widehat{\boldsymbol{\Lambda}}_{\text{AB}}\\ \widehat{\boldsymbol{\Lambda}}_{\text{AB}}^T\widehat{\boldsymbol{\Lambda}}_{\text{AA}}^{-1} & \widehat{\boldsymbol{\Lambda}}_{\text{AB}}^T\widehat{\boldsymbol{\Lambda}}_{\text{AA}}^{-1}\widehat{\boldsymbol{\Lambda}}_{\text{AB}}+\widehat{E}\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\}\end{bmatrix}^{-1} \tag{13}$$

and

$$\begin{aligned}\widehat{\text{Asy.Cov}}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})\big) &= \frac{2\boldsymbol{D}^{+}_{d_{\text{R}}}(\widehat{\boldsymbol{\Sigma}}\otimes\widehat{\boldsymbol{\Sigma}})\boldsymbol{D}^{+T}_{d_{\text{R}}}}{m}\\ &\quad+\frac{\widehat{\phi}}{mn}\Big(2\widehat{E}\big\{\boldsymbol{\Psi}_9(\boldsymbol{U})\big\}-4\widehat{E}\big\{\boldsymbol{\Psi}_8(\boldsymbol{U})\big\}+\widehat{\boldsymbol{\Phi}}^T\big[\widehat{E}\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\}\big]^{-1}\widehat{\boldsymbol{\Phi}}\Big).\end{aligned} \tag{14}$$

In the general quasi-likelihood situation, the most common choice for $\widehat{\phi}$ is the method of moments estimator and is often labelled the *Pearson* estimator. For ordinary likelihood settings, such as for Gaussian and Gamma responses, $\widehat{\phi}$ could instead be the maximum likelihood estimator.

Let $(\boldsymbol{\beta}^0)_k$ denote the kth entry of $\boldsymbol{\beta}^0$. Then approximate $100(1-\alpha)\%$ confidence intervals for $(\boldsymbol{\beta}^0)_k$ based on (13) are

$$(\widehat{\boldsymbol{\beta}})_k \pm \Phi^{-1}(1-\tfrac{1}{2}\alpha)\sqrt{\left\{\widehat{\text{Asy.Cov}}(\widehat{\boldsymbol{\beta}})\right\}_{kk}}, \quad 1 \le k \le d_{\text{F}}. \tag{15}$$

The confidence intervals in (15) are analogous to those given in Section 4 of Jiang *et al.* (2022). For $1 \le k \le d_{\text{R}}$, (15) provides second term improvements of the Jiang *et al.* (2022) confidence intervals. For $d_{\text{R}}+1 \le k \le d_{\text{F}}$ both sets of confidence intervals are identical.

Improved confidence intervals for the random effects covariance parameters can be constructed in a similar fashion based on (14).

4.2 Other Utilities

The second term improvements of (8) may also be applied to Wald hypothesis tests and sample size calculations. Optimal design is another possible utility, but would require second term improvements of the type of theory given in Section 5 of Jiang *et al.* (2022).

5 Numerical Results

We conducted a simulation exercise aimed at understanding potential practical impacts of second term improvements to generalized linear mixed model asymptotics. The results are presented in this section.

Our simulation exercise involved generation of data sets from the $d_{\text{F}} = 5$ and $d_{\text{R}} = 2$ logistic mixed model

$$\begin{aligned} &Y_{ij}|X_{1ij}, X_{2ij}, X_{3ij}, X_{4ij}, U_i \text{ independently distributed as}\\ &\quad \text{Bernoulli}\Big(1/\big(1+\exp[-\{\beta_0^0 + U_{0i} + (\beta_1^0 + U_{1i})X_{1ij} + \beta_2^0 X_{2ij} + \beta_3^0 X_{3ij} + \beta_4^0 X_{4ij}\}]\big)\Big), \\ &\text{where the } \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} \text{ are independent } N(\mathbf{0}, \boldsymbol{\Sigma}^0) \text{ random vectors}, \ 1 \le i \le m, \ 1 \le j \le n. \end{aligned} \tag{16}$$

The 'true' parameter values were set to

$$\big(\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \beta_4^0\big) = (0.35, 0.96, -0.47, 1.06, -1.31) \quad \text{and} \quad \boldsymbol{\Sigma}^0 = \begin{bmatrix} 0.56 & -0.34 \\ -0.34 & 0.89 \end{bmatrix} \tag{17}$$

and the predictor data were generated from independent Uniform distributions on the unit interval. To assess potential large sample improvements afforded by the two-term asymptotic covariance expressions at (8) we varied m over the set $\{100, 150, \ldots, 500\}$ and fixed n at $m/10$. For each (m, n) pair we then simulated 500 data sets according to (16) and (17) and obtained approximate 95% confidence intervals for all model parameters according to the approach described in Section 4 of Jiang *et al.* (2022) and the second term improvements described in Section 4.1 of this article. The requisite bivariate integrals were obtained using the function `hcubature()` within the R language package cubature (Balasubramanian *et al.*, 2023).

Note that the confidence intervals for β_0^0, β_1^0 and the entries of $\boldsymbol{\Sigma}^0$ differ according to the two approaches since the estimators of these parameters have order m^{-1} asymptotic variances. The confidence intervals for β_2^0, β_3^0 and β_4^0 are unaffected by the second term asymptotic improvements since their estimators have order $(mn)^{-1}$ asymptotic variances.

Figure 1 compares the empirical coverages of confidence intervals with advertised levels of 95% for the one-term asymptotic variances of Jiang *et al.* (2022) and the two-term asymptotic variances that arise from (8). In Figure 1 we only consider the parameters that are affected by second term improvement. The empirical coverages for the other parameters are provided in the supplement.

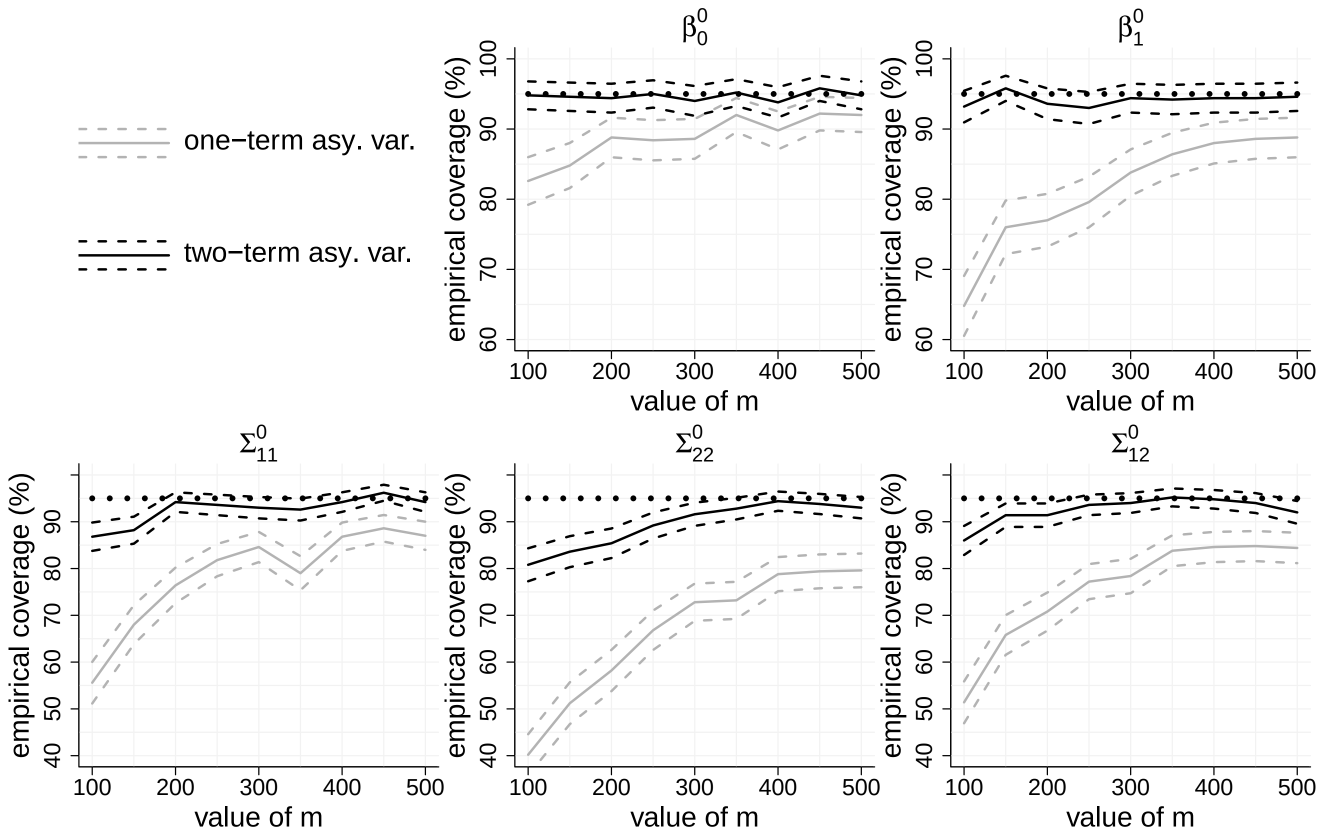

Figure 1: *Empirical coverage of confidence intervals from the simulation exercise described in the text. Each panel corresponds to a model parameter that is impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show, dependent on the number of groups m, the empirical coverage levels for confidence intervals that use both one-term and two-term asymptotic variance approximations. The dashed curves correspond to plus and minus two standard errors of the sample proportions. The within-group sample size, n, is fixed at $m/10$.*

It is clear from Figure 1 that our second term improvements lead to much better coverages for lower sample size situations. On the other hand, one-term confidence intervals are trivial to compute whilst the two-term versions require considerable computing involving numerical integration.

Simulation results such as those summarised by Figure 1 provide an appreciation for the practical trade-offs arising from precise asymptotics for generalised linear mixed models.

Acknowledgements

We are grateful to Alessandra Salvan and Nicola Sartori for advice related to this research. This research was supported by the Australian Research Council Discovery Project DP230101179.

References

Balasubramanian, N., Johnson, S.G., Hahn, T., Bouvier, A. & Kiêu, K. (2023). cubature 2.0.4.6: Adaptive multivariate integration over hypercubes. R package.
`https://r-project.org`

Bhaskaran, A. and Wand, M.P. (2023). Dispersion parameter extension of precise generalized linear mixed model asymptotics. *Statistics and Probability Letters*, **193**, Article 109691.

Jiang, J. & Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications, Second Edition.* New York: Springer.

Jiang, J., Wand, M.P. & Bhaskaran, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *Journal of the Royal Statistical Society, Series B*, **84**, 55–82.

McCulloch, C.E., Searle, S.R. & Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models. Second Edition.* New York: John Wiley & Sons.

Stroup, W.W. (2013). *Generalized Linear Mixed Models.* Boca Raton, Florida: CRC Press.

Supplement for:

Second Term Improvements to Generalised Linear Mixed Model Asymptotics

BY LUCA MAESTRINI[1], AISHWARYA BHASKARAN[2] AND MATT P. WAND[3]

[1]*The Australian National University,* [2]*Macquarie University and* [3]*University of Technology Sydney*

S.1 Introduction

The purpose of this supplement is to provide detailed derivational steps for the main result of Section 3.3 and further details on our simulation exercise. Sections S.2–S.5 provide relevant results concerning matrix algebra and multivariate calculus. In Sections S.6–S.9 we focus on the scores of the model parameters and their high-order asymptotic approximations. Sections S.10 and S.11 are concerned with approximation of the Fisher information matrix. The final stages of the derivations of (8) and (9) are given in Sections S.12 and S.14. Section S.15 provides some additional empirical coverage plots from the logistic mixed model simulation exercise described in Section 5.

S.2 Matrix Algebraic Results

The derivation of the results in Section 3.3 benefits from particular matrix results, which are summarized in this section.

For each $d \in \mathbb{N}$ the $d^2 \times \frac{1}{2}d(d+1)$ matrix $\boldsymbol{D}_d$ and $d^2 \times d^2$ matrix $\boldsymbol{K}_d$ are constant matrices containing zeroes and ones such that

$$\boldsymbol{D}_d\text{vech}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A}) \quad \text{for all symmetric } d \times d \text{ matrices } \boldsymbol{A}$$

and

$$\boldsymbol{K}_d\text{vec}(\boldsymbol{B}) = \text{vec}(\boldsymbol{B}^T) \quad \text{for all } d \times d \text{ matrices } \boldsymbol{B}.$$

Examples are

$$\boldsymbol{D}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{K}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The $\boldsymbol{D}_d$ are called *duplication* matrices, whilst the $\boldsymbol{K}_d$ are called *commutation* matrices. As stated in Section 2, the Moore-Penrose inverse of $\boldsymbol{D}_d$ is $\boldsymbol{D}_d^+ = (\boldsymbol{D}_d^T\boldsymbol{D}_d)^{-1}\boldsymbol{D}_d^T$. Chapter 3 of Magnus & Neudecker (1999) contains several results concerning these families of matrices, a few of which are relevant to the derivation of (8). For convenience, we list them here.

Theorem 9(c) in Chapter 3 of Magnus & Neudecker (1999) implies that for any $d \times d$ matrix $\boldsymbol{A}$ and $d \times 1$ vector $\boldsymbol{b}$, we have

$$\boldsymbol{K}_d(\boldsymbol{A} \otimes \boldsymbol{b}) = \boldsymbol{b} \otimes \boldsymbol{A}. \tag{S.1}$$

Theorem 12(a) in the same chapter asserts that

$$\boldsymbol{K}_d\boldsymbol{D}_d = \boldsymbol{D}_d \tag{S.2}$$

and implies that, for any $d \times d$ matrix $\boldsymbol{A}$,

$$\boldsymbol{D}_d^T\text{vec}(\boldsymbol{A}) = \boldsymbol{D}_d^T\text{vec}(\boldsymbol{A}^T). \tag{S.3}$$

Also, Theorem 13(b) and Theorem 13(d) provide for a $d \times d$ matrix $\boldsymbol{A}$

$$\boldsymbol{D}_d\boldsymbol{D}_d^+(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{D}_d^{+T} = (\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{D}_d^{+T} \tag{S.4}$$

and, assuming that $\boldsymbol{A}$ is invertible,

$$\left\{\boldsymbol{D}_d^T(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{D}_d\right\}^{-1}=\boldsymbol{D}_d^+(\boldsymbol{A}^{-1}\otimes\boldsymbol{A}^{-1})\boldsymbol{D}_d^{+T}. \tag{S.5}$$

Lastly, we state two matrix identities that are used in the derivations. For matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ such that $\boldsymbol{ABC}$ is defined, we have

$$\text{vec}(\boldsymbol{ABC})=(\boldsymbol{C}^T\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B}). \tag{S.6}$$

For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$, we have

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})=(\boldsymbol{AC})\otimes(\boldsymbol{BD}). \tag{S.7}$$

S.3 Multivariate Derivative Notation

For f a smooth real-valued function of the d-variate argument $\boldsymbol{x}\equiv(x_1,\ldots,x_d)$, let $\nabla f(\boldsymbol{x})$ denote the $d\times 1$ vector with rth entry $\partial f(\boldsymbol{x})/\partial x_r$, $\nabla^2 f(\boldsymbol{x})$ denote the $d\times d$ matrix with (r,s) entry $\partial^2 f(\boldsymbol{x})/(\partial x_r\partial x_s)$ and $\nabla^3 f(\boldsymbol{x})$ denote the $d\times d\times d$ array with (r,s,t) entry $\partial^3 f(\boldsymbol{x})/(\partial x_r\partial x_s\partial x_t)$.

S.4 Three-Term Taylor Series Expansion of Gradient Vectors

Consider $f:\mathbb{R}^d\to\mathbb{R}$. Then, for sufficiently smooth f, the three-term Taylor expansion of

$$f(\boldsymbol{x}+\boldsymbol{h})\quad\text{where }\boldsymbol{x}\equiv(x_1,\ldots,x_d)\text{ and }\boldsymbol{h}\equiv(h_1,\ldots,h_d)$$

is

$$f(\boldsymbol{x}+\boldsymbol{h})=f(\boldsymbol{x})+\sum_{r=1}^d\{\nabla f(\boldsymbol{x})\}_r h_r+\tfrac{1}{2}\sum_{r=1}^d\sum_{s=1}^d\{\nabla^2 f(\boldsymbol{x})\}_{rs}h_rh_s+\ldots. \tag{S.8}$$

Now consider $\alpha:\mathbb{R}^d\to\mathbb{R}$ and its gradient function $\nabla\alpha:\mathbb{R}^d\to\mathbb{R}^d$. If (S.8) is applied to each entry of $(\nabla\alpha)(\boldsymbol{x}+\boldsymbol{h})$ then we have

$$\begin{bmatrix}\{(\nabla\alpha)(\boldsymbol{x}+\boldsymbol{h})\}_1\\ \vdots\\ \{(\nabla\alpha)(\boldsymbol{x}+\boldsymbol{h})\}_d\end{bmatrix}=\begin{bmatrix}\{(\nabla\alpha)(\boldsymbol{x})\}_1+\sum_{r=1}^d\{\nabla^2\alpha(\boldsymbol{x})\}_{r1}h_r+\frac{1}{2}\sum_{r=1}^d\sum_{s=1}^d\{\nabla^3\alpha(\boldsymbol{x})\}_{rs1}h_rh_s\\ \vdots\\ \{(\nabla\alpha)(\boldsymbol{x})\}_d+\sum_{r=1}^d\{\nabla^2\alpha(\boldsymbol{x})\}_{rd}h_r+\frac{1}{2}\sum_{r=1}^d\sum_{s=1}^d\{\nabla^3\alpha(\boldsymbol{x})\}_{rsd}h_rh_s\end{bmatrix}+\ldots.$$

From this it is clear that

$$(\nabla\alpha)(\boldsymbol{x}+\boldsymbol{h})=(\nabla\alpha)(\boldsymbol{x})+\{(\nabla^2\alpha)(\boldsymbol{x})\}\boldsymbol{h}+\tfrac{1}{2}\{(\nabla^3\alpha)(\boldsymbol{x})\}\bigstar(\boldsymbol{hh}^T)+\ldots. \tag{S.9}$$

where the $\bigstar$ notation is as defined by (6).

S.5 Higher Order Approximation of Multivariate Integral Ratios

The main tool for approximation of the Fisher information matrix of (3) is higher order Laplace-type approximation of multivariate integral ratios. Appendix A of Miyata (2004) provides such a result, which states that for smooth real-valued d-variate functions g, c and h,

$$\begin{aligned}\frac{\int_{\mathbb{R}^d}g(\boldsymbol{x})c(\boldsymbol{x})\exp\{-nh(\boldsymbol{x})\}\,d\boldsymbol{x}}{\int_{\mathbb{R}^d}c(\boldsymbol{x})\exp\{-nh(\boldsymbol{x})\}\,d\boldsymbol{x}}&=g(\boldsymbol{x}^*)+\frac{\nabla g(\boldsymbol{x}^*)^T\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla c(\boldsymbol{x}^*)}{nc(\boldsymbol{x}^*)}\\&+\frac{\text{tr}[\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\nabla^2 g(\boldsymbol{x}^*)]}{2n}-\frac{\nabla g(\boldsymbol{x}^*)^T\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\left[\nabla^3 h(\boldsymbol{x}^*)\bigstar\{\nabla^2 h(\boldsymbol{x}^*)\}^{-1}\right]}{2n}+O(n^{-2})\end{aligned} \tag{S.10}$$

where

$$\boldsymbol{x}^*\equiv\underset{\boldsymbol{x}\in\mathbb{R}^d}{\text{argmin}}\,h(\boldsymbol{x}).$$

S.6 Exact Score Expressions

For $1 \leq i \leq m$, let $p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}$ denote the conditional density function, or probability mass function, of $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$. Then let

$$\boldsymbol{S}_{\mathrm{A}i} \equiv \nabla_{\boldsymbol{\beta}_{\mathrm{A}}} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i), \quad \boldsymbol{S}_{\mathrm{B}i} \equiv \nabla_{\boldsymbol{\beta}_{\mathrm{B}}} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)$$

and

$$\boldsymbol{S}_{\mathrm{C}i} \equiv \nabla_{\mathrm{vech}(\boldsymbol{\Sigma})} \log p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{Y}_i|\boldsymbol{X}_i)$$

denote the ith contribution to the scores with respect to each of $\boldsymbol{\beta}_{\mathrm{A}}$, $\boldsymbol{\beta}_{\mathrm{B}}$ and vech($\boldsymbol{\Sigma}$). Then it is straightforward to show that the exact scores are

$$\boldsymbol{S}_{\mathrm{A}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} \boldsymbol{g}_{iA}(\boldsymbol{u}) c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}}, \tag{S.11}$$

$$\boldsymbol{S}_{\mathrm{B}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} \boldsymbol{g}_{iB}(\boldsymbol{u}) c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}} \tag{S.12}$$

and

$$\boldsymbol{S}_{\mathrm{C}i} = \frac{\int_{\mathbb{R}^{d_{\mathrm{R}}}} \boldsymbol{g}_{iC}(\boldsymbol{u}) c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}}{\int_{\mathbb{R}^{d_{\mathrm{R}}}} c_S(\boldsymbol{u}) \exp\{-nh_i(\boldsymbol{u})\}\, d\boldsymbol{u}} - \tfrac{1}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T \mathrm{vec}(\boldsymbol{\Sigma}^{-1}) \tag{S.13}$$

where

$$\begin{aligned}
& c_S(\boldsymbol{u}) \equiv \exp(-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}), \quad \boldsymbol{g}_{iA}(\boldsymbol{u}) \equiv \boldsymbol{\Sigma}^{-1}\boldsymbol{u},\\
& \boldsymbol{g}_{iB}(\boldsymbol{u}) \equiv \frac{1}{\phi}\sum_{j=1}^{n_i} \boldsymbol{X}_{\mathrm{B}ij}\{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\},\\
& \boldsymbol{g}_{iC}(\boldsymbol{u}) \equiv \tfrac{1}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{u}\boldsymbol{u}^T)\\
\text{and}\quad & h_i(\boldsymbol{u}) \equiv -\frac{1}{n\phi}\sum_{j=1}^{n_i}\big\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij} - b\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\big\}.
\end{aligned}$$

An integration by parts step is used to obtain the $\boldsymbol{S}_{\mathrm{A}i}$ expression.

In the upcoming sections we obtain asymptotic approximations of $\boldsymbol{S}_{\mathrm{A}i}$ $\boldsymbol{S}_{\mathrm{B}i}$ and $\boldsymbol{S}_{\mathrm{C}i}$. Key quantities for these approximations are

$$\boldsymbol{U}_i^* \equiv \underset{\boldsymbol{u}\in\mathbb{R}^{d_{\mathrm{R}}}}{\mathrm{argmin}}\, h_i(\boldsymbol{u}), \quad 1 \leq i \leq m.$$

S.7 Definitions of Key Summation Quantities

Our derivation of (8) involves manipulations of particular summation quantities, which are defined in this section. At the end of this section we state some important moment-type relationships between the quantities.

For each $1 \leq i \leq m$, define $\mathcal{G}_{\mathrm{A}i}$, $\mathcal{G}_{\mathrm{B}i}$, $\mathcal{H}_{\mathrm{AA}i}$, $\mathcal{H}_{\mathrm{AB}i}$, and $\mathcal{H}_{\mathrm{BB}i}$ as follows:

$$\mathcal{G}_{\mathrm{A}i} \equiv \sum_{j=1}^{n_i}\{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\}\boldsymbol{X}_{\mathrm{A}ij},$$

$$\mathcal{G}_{\mathrm{B}i} \equiv \sum_{j=1}^{n_i}\{Y_{ij} - b'\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\}\boldsymbol{X}_{\mathrm{B}ij},$$

$$\mathcal{H}_{\mathrm{AA}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T,$$

$$\mathcal{H}_{\text{AB}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\big)\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{B}ij}^T$$

$$\text{and}\quad \mathcal{H}_{\text{BB}i} \equiv \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\big)\boldsymbol{X}_{\text{B}ij}\boldsymbol{X}_{\text{B}ij}^T.$$

In a similar vein, define $\mathcal{H}'_{\text{AAA}i}$ to be the $d_{\text{R}}\times d_{\text{R}}\times d_{\text{R}}$ array with (r,s,t) entry equal to

$$\sum_{j=1}^{n_i} b'''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\big)(\boldsymbol{X}_{\text{A}ij})_r(\boldsymbol{X}_{\text{A}ij})_s(\boldsymbol{X}_{\text{A}ij})_t$$

and $\mathcal{H}'_{\text{AAB}i}$ to be the $d_{\text{R}}\times d_{\text{R}}\times d_{\text{B}}$ array with (r,s,t) entry equal to

$$\sum_{j=1}^{n_i} b'''\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\big)(\boldsymbol{X}_{\text{A}ij})_r(\boldsymbol{X}_{\text{A}ij})_s(\boldsymbol{X}_{\text{B}ij})_t$$

where

$$d_{\text{B}} \equiv d_{\text{F}} - d_{\text{R}}.$$

The following relationships are of fundamental importance for the derivation of (8):

$$\begin{gathered} E\big(\mathcal{G}_{\text{A}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\big)=\mathbf{0},\quad E\big(\mathcal{G}_{\text{B}i}|\boldsymbol{X}_i,\boldsymbol{U}_i\big)=\mathbf{0},\\ E\big(\mathcal{G}_{\text{A}i}^{\otimes 2}|\boldsymbol{X}_i,\boldsymbol{U}_i\big)=\phi\mathcal{H}_{\text{AA}i},\quad E\big(\mathcal{G}_{\text{A}i}\mathcal{G}_{\text{B}i}^T|\boldsymbol{X}_i,\boldsymbol{U}_i\big)=\phi\mathcal{H}_{\text{AB}i}\quad\text{and}\quad E\big(\mathcal{G}_{\text{B}i}^{\otimes 2}|\boldsymbol{X}_i,\boldsymbol{U}_i\big)=\phi\mathcal{H}_{\text{BB}i}\end{gathered} \tag{S.14}$$

where, throughout this supplement,

$$\boldsymbol{v}^{\otimes 2}\equiv \boldsymbol{v}\boldsymbol{v}^T \quad \text{for any column vector } \boldsymbol{v}.$$

Also note that

$$\begin{gathered}\mathcal{G}_{\text{A}i}=O_p(n^{1/2})\mathbf{1}_{d_{\text{R}}},\quad \mathcal{G}_{\text{B}i}=O_p(n^{1/2})\mathbf{1}_{d_{\text{B}}},\quad \mathcal{H}_{\text{AA}i}=O_p(n)\mathbf{1}_{d_{\text{R}}}^{\otimes 2},\\ \mathcal{H}_{\text{BB}i}=O_p(n)\mathbf{1}_{d_{\text{B}}}^{\otimes 2},\quad \mathcal{H}_{\text{AB}i}=O_p(n)\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{B}}}^T\end{gathered}$$

and that all entries of $\mathcal{H}'_{\text{AAA}i}$ and $\mathcal{H}'_{\text{AAB}i}$ are $O_p(n)$.

S.8 Approximation of $\boldsymbol{U}_i^*$

Use of (S.10) to approximate $\boldsymbol{S}_{\text{A}i}$, $\boldsymbol{S}_{\text{B}i}$ and $\boldsymbol{S}_{\text{C}i}$ requires approximation of $\boldsymbol{U}_i^*$. Introduce the notation $\mathcal{C}_i(\boldsymbol{u})\equiv n\phi h_i(\boldsymbol{u})$. Then $\boldsymbol{U}_i^*$ satisfies

$$\nabla\mathcal{C}_i(\boldsymbol{U}_i^*)=\mathbf{0}$$

where

$$\nabla\mathcal{C}_i(\boldsymbol{u})\equiv -\sum_{j=1}^{n_i}\Big\{Y_{ij}-b'\big((\boldsymbol{\beta}_{\text{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij}+\boldsymbol{\beta}_{\text{B}}^T\boldsymbol{X}_{\text{B}ij}\big)\Big\}\boldsymbol{X}_{\text{A}ij}.$$

Then, from (S.9) we have

$$\begin{aligned}\nabla\mathcal{C}_i(\boldsymbol{U}_i^*) &= \nabla\mathcal{C}_i(\boldsymbol{U}_i+\boldsymbol{U}_i^*-\boldsymbol{U}_i)\\ &= \nabla\mathcal{C}_i(\boldsymbol{U}_i)+\{\nabla^2\mathcal{C}_i(\boldsymbol{U}_i)\}(\boldsymbol{U}_i^*-\boldsymbol{U}_i)\\ &\qquad +\tfrac{1}{2}\left\{\nabla^3\mathcal{C}_i(\boldsymbol{U}_i)\right\}\bigstar\{(\boldsymbol{U}_i^*-\boldsymbol{U}_i)(\boldsymbol{U}_i^*-\boldsymbol{U}_i)^T\}+\ldots.\end{aligned}$$

Next we seek explicit expressions for $\nabla^2\mathcal{C}_i(\boldsymbol{u})$ and $\nabla^3\mathcal{C}_i(\boldsymbol{u})$. Standard vector calculus arguments lead to

$$\begin{aligned}\nabla^2\mathcal{C}_i(\boldsymbol{u}) &= \sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\\ &= \left[\sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)(\boldsymbol{X}_{\mathrm{A}ij})_r(\boldsymbol{X}_{\mathrm{A}ij})_s\right]_{1\le r,s\le d_{\mathrm{R}}}.\end{aligned}$$

Then, the three-dimension array of all third order partial derivatives of $\mathcal{C}_i(u)$ is

$$\nabla^3\mathcal{C}_i(\boldsymbol{u}) = \left[\sum_{j=1}^{n_i} b'''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)(\boldsymbol{X}_{\mathrm{A}ij})_r(\boldsymbol{X}_{\mathrm{A}ij})_s(\boldsymbol{X}_{\mathrm{A}ij})_t\right]_{1\le r,s,t\le d_{\mathrm{R}}}.$$

We then have

$$\nabla\mathcal{C}_i(\boldsymbol{U}_i^*) = -\mathcal{G}_{\mathrm{A}i}+\mathcal{H}_{\mathrm{AA}i}(\boldsymbol{U}_i^*-\boldsymbol{U}_i)+\tfrac{1}{2}\mathcal{H}'_{\mathrm{AAA}i}\bigstar\{(\boldsymbol{U}_i^*-\boldsymbol{U}_i)(\boldsymbol{U}_i^*-\boldsymbol{U}_i)^T\}+\dots$$

and so $\nabla\mathcal{C}_i(\boldsymbol{U}_i^*)=\mathbf{0}$ is equivalent to

$$\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} = (\boldsymbol{U}_i^*-\boldsymbol{U}_i)+\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big[\mathcal{H}'_{\mathrm{AAA}i}\bigstar\{(\boldsymbol{U}_i^*-\boldsymbol{U}_i)(\boldsymbol{U}_i^*-\boldsymbol{U}_i)^T\}\Big]+\dots. \tag{S.15}$$

We now invert (S.15) using the set-up given around equations (9.43) and (9.44) of Pace & Salvan (1997). To match the notation given there, set

$$\boldsymbol{y}\equiv\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\quad\text{and}\quad\boldsymbol{x}\equiv\boldsymbol{U}_i^*-\boldsymbol{U}_i.$$

Then, in keeping with the displayed equation just before (9.43) of Pace & Salvan (1997) and using their superscript and subscript conventions, we have

$$y^a\equiv\text{the }a\text{th entry of }\boldsymbol{y}\quad\text{and}\quad x^a\equiv\text{the }a\text{th entry of }\boldsymbol{x}.$$

Also,

$$x^{rs}\equiv x^rx^s=\text{the }(r,s)\text{ entry of }\boldsymbol{x}\boldsymbol{x}^T=\text{the }(r,s)\text{ entry of }(\boldsymbol{U}_i^*-\boldsymbol{U}_i)^{\otimes 2}.$$

Then

$$y^a=x^a+A^a_{rs}x^{rs}+\dots.$$

where

$$\begin{aligned}A^a_{rs}\boldsymbol{x}^{rs}&=\text{the }a\text{th entry of }\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big[\mathcal{H}'_{\mathrm{AAA}i}\bigstar\{(\boldsymbol{U}_i^*-\boldsymbol{U}_i)(\boldsymbol{U}_i^*-\boldsymbol{U}_i)^T\}\Big]\\ &=\text{the }a\text{th entry of }\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar(\boldsymbol{x}\boldsymbol{x}^T)\Big\}.\end{aligned}$$

From equations (9.43) and (9.44) of Pace & Salvan (1997),

$$\begin{aligned}x^a&=y^a-A^a_{rs}y^{rs}+\dots\\ &=y^a-\text{the }a\text{th entry of }\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar(\boldsymbol{y}\boldsymbol{y}^T)\Big\}+\dots\\ &=y^a-\text{the }a\text{th entry of }\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\Big)^{\otimes 2}\Big\}+\dots.\end{aligned}$$

This results in the following three-term approximation of $\boldsymbol{U}_i^*$:

$$\boldsymbol{U}_i^*=\boldsymbol{U}_i+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}-\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}+O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}}. \tag{S.16}$$

S.9 Score Asymptotic Approximation

We are now ready to obtain approximations of the scores $\boldsymbol{S}_{\mathrm{A}i}$, $\boldsymbol{S}_{\mathrm{B}i}$ and $\boldsymbol{S}_{\mathrm{C}i}$ with accuracies that are sufficient for the two-term asymptotic covariance matrices of (8).

S.9.1 Approximation of $\boldsymbol{S}_{\mathrm{A}i}$

For each $1 \le r \le d_{\mathrm{R}}$, let $\boldsymbol{e}_r$ denote the $d_{\mathrm{R}} \times 1$ vector having rth entry equal to 1 and zeroes elsewhere.

S.9.1.1 The (S.10) First Term Contribution

For each $1 \le r \le d_{\mathrm{R}}$, the contribution to the rth entry of $\boldsymbol{S}_{\mathrm{A}i}$ from the first term on the right-hand side of (S.10) is the rth entry of $\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*$. In view of (S.16) we obtain the following contribution to $\boldsymbol{S}_{\mathrm{A}i}$:

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + \boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \tfrac{1}{2}\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} + O_p(n^{-3/2})\boldsymbol{1}_{d_{\mathrm{R}}}.$$

S.9.1.2 The (S.10) Second Term Contribution

Noting that

$$\nabla\{\boldsymbol{e}_r^T\boldsymbol{g}_{iA}(\boldsymbol{u})\} = \boldsymbol{e}_r^T\boldsymbol{\Sigma}^{-1} \quad \text{and} \quad \nabla c_S(\boldsymbol{u}) = -c_S(\boldsymbol{u})\boldsymbol{\Sigma}^{-1}\boldsymbol{u},$$

the contribution to the rth entry of $\boldsymbol{S}_{\mathrm{A}i}$ from the second term on the right-hand side of (S.10) is

$$-\phi\boldsymbol{e}_r^T\boldsymbol{\Sigma}^{-1}\left\{\sum_{j=1}^{n_i} b''\big((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i^*)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right\}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*. \tag{S.17}$$

Substitution of (S.16) into (S.17) then leads to the following contribution to $\boldsymbol{S}_{\mathrm{A}i}$:

$$-\phi\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_p(n^{-3/2})\boldsymbol{1}_{d_{\mathrm{R}}}.$$

S.9.1.3 The (S.10) Third Term Contribution

Noting that $\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iA}(\boldsymbol{u})\} = \boldsymbol{O}$, the contribution to $\boldsymbol{S}_{\mathrm{A}i}$ from the third term on the right-hand side of (S.10) is $\boldsymbol{0}$.

S.9.1.4 The (S.10) Fourth Term Contribution

Via arguments similar to those given in Section S.9.1.2, the contribution to $\boldsymbol{S}_{\mathrm{A}i}$ from the fourth term on the right-hand side of (S.10) is

$$-\frac{\phi}{2}\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big) + O_p(n^{-3/2})\boldsymbol{1}_{d_{\mathrm{R}}}.$$

S.9.1.5 The Resultant Score Approximation

On combining the results of Sections S.9.1.1–S.9.1.4, we obtain

$$\begin{aligned}\boldsymbol{S}_{\mathrm{A}i} = {}&\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + \boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\\ &-\phi\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i - \frac{\phi}{2}\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big) + O_p(n^{-3/2})\boldsymbol{1}_{d_{\mathrm{R}}}.\end{aligned} \tag{S.18}$$

S.9.2 Approximation of $\boldsymbol{S}_{\mathrm{B}i}$

For each $1 \le r \le d_{\mathrm{R}}$, let $\boldsymbol{e}_r$ denote the $d_{\mathrm{B}} \times 1$ vector having rth entry equal to 1 and zeroes elsewhere.

S.9.2.1 The (S.10) First Term Contribution

The contribution to $\boldsymbol{S}_{\text{B}i}$ from the first term on the right-hand side of (S.10) is

$$\boldsymbol{g}_{iB}(\boldsymbol{U}_i^*) = \frac{1}{\phi}\sum_{j=1}^{n_i}\boldsymbol{X}_{\text{B}ij}\big\{Y_{ij} - b'\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i^*)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\big\}. \tag{S.19}$$

Next note that, with (S.16) as a basis,

$$\begin{aligned}
&b'\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i^*)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\\
&\quad = b'\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\\
&\qquad + \boldsymbol{X}_{\text{A}ij}^T(\boldsymbol{U}_i^* - \boldsymbol{U}_i)b''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\\
&\qquad + \frac{1}{2}\boldsymbol{X}_{\text{A}ij}^T(\boldsymbol{U}_i^* - \boldsymbol{U}_i)^{\otimes 2}\boldsymbol{X}_{\text{A}ij}b'''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big) + O_p(n^{-3/2})\\
&\quad = b'\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\\
&\qquad + \boldsymbol{X}_{\text{A}ij}^T\left[\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i} - \frac{1}{2}\mathcal{H}_{\text{AA}i}^{-1}\left\{\mathcal{H}'_{\text{AAA}i}\bigstar\left(\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\mathcal{G}_{\text{A}i}^T\mathcal{H}_{\text{AA}i}^{-1}\right)\right\}\right]\\
&\qquad\qquad \times b''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\\
&\qquad + \frac{1}{2}\boldsymbol{X}_{\text{A}ij}^T\left(\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\right)^{\otimes 2}\boldsymbol{X}_{\text{A}ij}b'''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big) + O_p(n^{-3/2}).
\end{aligned}$$

Substitution of this result into (S.19) leads to the first term of $\boldsymbol{S}_{\text{B}i}$ equalling

$$\begin{aligned}
&\frac{1}{\phi}\left(\mathcal{G}_{\text{B}i} - \mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\right) + \frac{1}{2\phi}\mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\left\{\mathcal{H}'_{\text{AAA}i}\bigstar\left(\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\mathcal{G}_{\text{A}i}^T\mathcal{H}_{\text{AA}i}^{-1}\right)\right\}\\
&\qquad - \frac{1}{2\phi}\left\{\mathcal{H}'_{\text{AAB}i}\bigstar\left(\mathcal{H}_{\text{AA}i}^{-1}\mathcal{G}_{\text{A}i}\mathcal{G}_{\text{A}i}^T\mathcal{H}_{\text{AA}i}^{-1}\right)\right\} + O_p(n^{-1/2})\mathbf{1}_{d_\text{R}}.
\end{aligned}$$

S.9.2.2 The (S.10) Second Term Contribution

Noting that

$$\nabla\{\boldsymbol{e}_r^T\boldsymbol{g}_{iB}(\boldsymbol{u})\} = -\frac{1}{\phi}\sum_{j=1}^{n_i}b''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{u})^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\boldsymbol{e}_r^T\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{B}ij}^T \tag{S.20}$$

and recalling that $\nabla c_S(\boldsymbol{u}) = -c_S(\boldsymbol{u})\boldsymbol{\Sigma}^{-1}\boldsymbol{u}$, the contribution from the second term on the right-hand side of (S.10) to $\boldsymbol{S}_{\text{B}i}$ is

$$\begin{aligned}
&\left\{\sum_{j=1}^{n_i}b''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i^*)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{B}ij}^T\right\}^T\\
&\qquad \times\left\{\sum_{j=1}^{n_i}b''\big((\boldsymbol{\beta}_\text{A} + \boldsymbol{U}_i^*)^T\boldsymbol{X}_{\text{A}ij} + \boldsymbol{\beta}_\text{B}^T\boldsymbol{X}_{\text{B}ij}\big)\boldsymbol{X}_{\text{A}ij}\boldsymbol{X}_{\text{A}ij}^T\right\}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*.
\end{aligned}$$

Substitution of (S.16) then leads to the contribution to $\boldsymbol{S}_{\text{B}i}$ from the second term of (S.10) equalling

$$\mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_p(n^{-1/2})\mathbf{1}_{d_\text{R}}.$$

S.9.2.3 The (S.10) Third Term Contribution

The rth entry of the contribution to $\boldsymbol{S}_{\mathrm{B}i}$ from the third term of (S.10) is

$$\frac{1}{2n}\sum_{s=1}^{d_{\mathrm{B}}}\sum_{t=1}^{d_{\mathrm{B}}}\big[\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iB}(\boldsymbol{U}_i^*)\}\big]_{st}\big[\{\nabla^2 h_i(\boldsymbol{U}_i^*)\}^{-1}\big]_{st}$$

$$=\frac{\phi}{2}\sum_{s=1}^{d_{\mathrm{B}}}\sum_{t=1}^{d_{\mathrm{B}}}\big[\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iB}(\boldsymbol{U}_i)\}\big]_{st}\big(\mathcal{H}_{\mathrm{AA}i}^{-1}\big)_{st}+O_p(n^{-1/2}).$$

However, the (s,t) entry of $\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iB}(\boldsymbol{U}_i)\}$ is

$$-\frac{1}{\phi}\sum_{j=1}^{n_i}b'''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)(\boldsymbol{e}_r^T\boldsymbol{X}_{\mathrm{A}ij})(\boldsymbol{e}_s^T\boldsymbol{X}_{\mathrm{A}ij})(\boldsymbol{e}_t^T\boldsymbol{X}_{\mathrm{B}ij})=-\frac{1}{\phi}\big(\mathcal{H}'_{\mathrm{AAB}i}\big)_{rst}.$$

Noting (6), the contribution to $\boldsymbol{S}_{\mathrm{B}i}$ from the third term of (S.10) is

$$-\frac{1}{2}\mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}+O_p(n^{-1/2})\mathbf{1}_{d_{\mathrm{R}}}.$$

S.9.2.4 The (S.10) Fourth Term Contribution

With the aid of (S.20), the contribution to $\boldsymbol{S}_{\mathrm{B}i}$ from the fourth term of (S.10) is

$$\frac{1}{2n\phi}\left\{\sum_{j=1}^{n_i}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i^*)^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{B}ij}^T\right\}^T\{\nabla^2 h_i(\boldsymbol{U}_i^*)\}^{-1}\left[\nabla^3 h_i(\boldsymbol{U}_i^*)\bigstar\{\nabla^2 h_i(\boldsymbol{U}_i^*)\}^{-1}\right]$$

$$=\frac{1}{2}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)+O_p(n^{-1/2})\mathbf{1}_{d_{\mathrm{R}}}.$$

S.9.2.5 The Resultant Score Approximation

On combining each of the contributions, we obtain

$$\begin{aligned}\boldsymbol{S}_{\mathrm{B}i}=&\frac{1}{\phi}\Big(\mathcal{G}_{\mathrm{B}i}-\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\Big)+\frac{1}{2\phi}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\\&-\frac{1}{2\phi}\Big\{\mathcal{H}'_{\mathrm{AAB}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}+\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\\&-\frac{1}{2}\mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}+\frac{1}{2}\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)+O_p(n^{-1/2})\mathbf{1}_{d_{\mathrm{R}}}.\end{aligned}\tag{S.21}$$

S.9.3 Approximation of $\boldsymbol{S}_{\mathrm{C}i}$

For each $1\le r\le\frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$ let $\boldsymbol{e}_r$ denote the $d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2\times 1$ vector with 1 in the rth position and zeroes elsewhere.

S.9.3.1 The (S.10) First Term Contribution

For each $1\le r\le\frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$, the rth entry of the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the first term of (S.10) is

$$\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{U}_i^*)=\frac{1}{2}\boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}\big((\boldsymbol{U}_i^*)^{\otimes 2}\big).$$

Since

$$
\begin{aligned}
(\boldsymbol{U}_i^*)^{\otimes 2} &= \Big[\boldsymbol{U}_i + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i} - \tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\} + O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}}\Big]^{\otimes 2} \\
&= \boldsymbol{U}_i^{\otimes 2} + \boldsymbol{U}_i\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\quad -\tfrac{1}{2}\boldsymbol{U}_i\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\quad -\tfrac{1}{2}\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\boldsymbol{U}_i^T + O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}}^{\otimes 2},
\end{aligned}
$$

and noting (S.3) and (S.6), the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the first term of (S.10) is

$$
\begin{aligned}
&\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\boldsymbol{\Sigma}^{-1}\Big[\boldsymbol{U}_i\boldsymbol{U}_i^T + 2\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T + \mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1} \\
&\qquad\qquad -\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\boldsymbol{U}_i^T\Big]\boldsymbol{\Sigma}^{-1}\Big) + O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2.}
\end{aligned}
$$

S.9.3.2 The (S.10) Second Term Contribution

Noting that, for each $1 \le r \le d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2$,

$$
\big[\nabla\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{u})\}\big]^T = \frac{1}{2}\boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(\boldsymbol{u}\otimes\boldsymbol{I}) + (\boldsymbol{I}\otimes\boldsymbol{u})\big\} \tag{S.22}
$$

and keeping in mind that $\nabla c_S(\boldsymbol{u}) = -c_S(\boldsymbol{u})\boldsymbol{\Sigma}^{-1}\boldsymbol{u}$, the contribution from the second term on the right-hand side of (S.10) to $\boldsymbol{S}_{\mathrm{C}i}$ is

$$
\begin{aligned}
&-\frac{\phi}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(\boldsymbol{U}_i^*\otimes\boldsymbol{I}) + (\boldsymbol{I}\otimes\boldsymbol{U}_i^*)\big\} \\
&\qquad\times\left\{\sum_{j=1}^{n_i}b''\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{U}_i^*)^T\boldsymbol{X}_{\mathrm{A}ij} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\boldsymbol{X}_{\mathrm{A}ij}\boldsymbol{X}_{\mathrm{A}ij}^T\right\}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^* \\
&= -\frac{\phi}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(\boldsymbol{U}_i\otimes\boldsymbol{I}) + (\boldsymbol{I}\otimes\boldsymbol{U}_i)\big\}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i + O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2} \\
&= -\phi\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\Big) + O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}.
\end{aligned}
$$

The last step makes use of (S.1), (S.3) and (S.6).

S.9.3.3 The (S.10) Third Term Contribution

The derivation of the (S.10) third term contribution to $\boldsymbol{S}_{\mathrm{C}i}$ benefits from notation and a result concerning the inverse of the vec operator. For $d \in \mathbb{N}$, if $\boldsymbol{b}$ is a $d^2 \times 1$ vector then $\mathrm{vec}^{-1}(\boldsymbol{b})$ is the $d \times d$ matrix such that $\mathrm{vec}\big(\mathrm{vec}^{-1}(\boldsymbol{b})\big) = \boldsymbol{b}$.

Lemma 1. *Let $d \in \mathbb{N}$, $\boldsymbol{a}$ be a $d \times 1$ vector and $\boldsymbol{b}$ be a $d^2 \times 1$ vector. Then*

$$
(\boldsymbol{a}^T\otimes\boldsymbol{I})\boldsymbol{b} = \mathrm{vec}^{-1}(\boldsymbol{b})\boldsymbol{a} \quad \textit{and} \quad (\boldsymbol{I}\otimes\boldsymbol{a}^T)\boldsymbol{b} = \mathrm{vec}^{-1}(\boldsymbol{b})^T\boldsymbol{a}.
$$

Lemma 1 is a relatively simple consequence of (S.6). To prove the first part of Lemma 1, note that its right-hand side is

$$
\mathrm{vec}^{-1}(\boldsymbol{b})\boldsymbol{a} = \mathrm{vec}\big(\mathrm{vec}^{-1}(\boldsymbol{b})\boldsymbol{a}\big) = \mathrm{vec}\big(\boldsymbol{I}\,\mathrm{vec}^{-1}(\boldsymbol{b})\boldsymbol{a}\big) = (\boldsymbol{a}^T\otimes\boldsymbol{I})\mathrm{vec}\big(\mathrm{vec}^{-1}(\boldsymbol{b})\big) = (\boldsymbol{a}^T\otimes\boldsymbol{I})\boldsymbol{b}.
$$

The proof of the second part of Lemma 1 is similar.

For each $1 \le r \le \frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$, the rth entry of the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the third term of (S.10) is

$$\frac{1}{2n}\mathrm{tr}\left[\{\nabla^2 h_i(\boldsymbol{U}_i)\}^{-1}\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{U}_i^*)\}\right].$$

Next note from (S.22) that

$$d\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{u})\} = \frac{1}{2}\boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(\boldsymbol{u}\otimes\boldsymbol{I})+(\boldsymbol{I}\otimes\boldsymbol{u})\big\}d\boldsymbol{u}.$$

Using Lemma 1 we then have

$$\begin{aligned}2d^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{u})\} &= \boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(d\boldsymbol{u}\otimes\boldsymbol{I})+(\boldsymbol{I}\otimes d\boldsymbol{u})\big\}d\boldsymbol{u}\\
&= \Big[\big\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\big\}^T(d\boldsymbol{u}\otimes\boldsymbol{I})+\big\{(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\big\}^T(\boldsymbol{I}\otimes d\boldsymbol{u})\Big]d\boldsymbol{u}\\
&= (d\boldsymbol{u})^T\Big[\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)+\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)^T\Big]d\boldsymbol{u}.\end{aligned}$$

From the second identification theorem of matrix differential calculus (e.g. Magnus & Neudecker, 1999) we then have

$$\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{u})\} = \frac{1}{2}\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)+\frac{1}{2}\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)^T$$

which does not depend on $\boldsymbol{u}$. Therefore $\nabla^2\boldsymbol{g}_k(\boldsymbol{U}_i^*)$ is a symmetric matrix that depends only on $\boldsymbol{\Sigma}$, which we denote as follows:

$$\boldsymbol{Q}(\boldsymbol{\Sigma};r) \equiv \frac{1}{2}\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)+\frac{1}{2}\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)^T.$$

Now note that

$$\begin{aligned}\frac{1}{2n}\mathrm{tr}\left[\{\nabla^2 h_i(\boldsymbol{U}_i^*)\}^{-1}\nabla^2\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{U}_i^*)\}\right] &= \mathrm{tr}\left[\{\nabla^2 h({}_i\boldsymbol{U}_i^*)\}^{-1}\boldsymbol{Q}(\boldsymbol{\Sigma};r)\right]\\
&= \frac{1}{2n}\mathrm{tr}\left[\{\nabla^2 h_i(\boldsymbol{U}_i)\}^{-1}\boldsymbol{Q}(\boldsymbol{\Sigma};r)\right]+O_p(n^{-3/2})\\
&= \frac{\phi}{2}\mathrm{tr}\left\{\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{Q}(\boldsymbol{\Sigma};r)\right\}+O_p(n^{-3/2}).\end{aligned}$$

The rth entry of the leading term of the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the third term on the right-hand side of (S.10) is

$$\begin{aligned}&\frac{\phi}{2}\mathrm{tr}\left\{\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{Q}(\boldsymbol{\Sigma};r)\right\}\\
&\quad= \frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T\mathrm{vec}\Big(\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)\Big)\\
&\qquad+\frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T\mathrm{vec}\Big(\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)^T\Big)\\
&\quad= \frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T\mathrm{vec}\Big(\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)\Big)\\
&\qquad+\frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T\boldsymbol{K}_{d_{\mathrm{R}}}\mathrm{vec}\Big(\mathrm{vec}^{-1}\Big((\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\Big)\Big)\\
&\quad= \frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r+\frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T\boldsymbol{K}_{d_{\mathrm{R}}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\\
&\quad= \frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r+\frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{K}_{d_{\mathrm{R}}}\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\\
&\quad= \frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r+\frac{\phi}{4}\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1})^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}}\boldsymbol{e}_r\\
&\quad= \frac{\phi}{2}\boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\mathcal{H}_{\mathrm{AA}i}^{-1}) = \frac{\phi}{2}\boldsymbol{e}_r^T\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}).\end{aligned}$$

Hence, contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the third term on the right-hand side of (S.10) is

$$\frac{\phi}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\big)+O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}.$$

S.9.3.4 The (S.10) Fourth Term Contribution

For each $1\le r\le\frac{1}{2}d_{\mathrm{R}}(d_{\mathrm{R}}+1)$, the rth entry of the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the fourth term on the right-hand side of (S.10) is

$$-\frac{1}{2n}\big[\nabla\{\boldsymbol{e}_r^T\boldsymbol{g}_{iC}(\boldsymbol{U}_i^*)\}\big]^T\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big[\nabla^3h_i(\boldsymbol{U}_i^*)\bigstar\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big].$$

Noting (S.22) and using (S.7), it follows that the contribution to $\boldsymbol{S}_{\mathrm{C}i}$ from the fourth term on the right-hand side of (S.10) is

$$\begin{aligned}
&-\frac{1}{4n}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big\{(\boldsymbol{U}_i^*\otimes\boldsymbol{I})+(\boldsymbol{I}\otimes\boldsymbol{U}_i^*)\big\}\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big[\nabla^3h_i(\boldsymbol{U}_i^*)\bigstar\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big]\\
&=-\frac{1}{4n}\boldsymbol{D}_{d_{\mathrm{R}}}^T\big[\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*)\otimes\boldsymbol{\Sigma}^{-1}\}+\{\boldsymbol{\Sigma}^{-1}\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i^*)\}\big]\\
&\qquad\qquad\times\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big[\nabla^3h_i(\boldsymbol{U}_i^*)\bigstar\{\nabla^2h_i(\boldsymbol{U}_i^*)\}^{-1}\big]\\
&=-\frac{\phi}{4}\boldsymbol{D}_{d_{\mathrm{R}}}^T\big[\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i)\otimes\boldsymbol{\Sigma}^{-1}\}+\{\boldsymbol{\Sigma}^{-1}\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i)\}\big]\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)+O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}\\
&=-\frac{\phi}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\boldsymbol{\Sigma}^{-1}\Big\{\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\boldsymbol{U}_i^T\Big\}\boldsymbol{\Sigma}^{-1}\Big)+O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}
\end{aligned}$$

where the last step follows from application of (S.3) and (S.6).

S.9.3.5 The Resultant Score Approximation

The resultant approximation of $\boldsymbol{S}_{\mathrm{C}i}$ is

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{C}i}=\frac{1}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T\mathrm{vec}\Big(\boldsymbol{\Sigma}^{-1}\Big[&\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma}+2\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\boldsymbol{U}_i^T+\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\\
&+\phi\mathcal{H}_{\mathrm{AA}i}^{-1}-2\phi\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\\
&-\mathcal{H}_{\mathrm{AA}i}^{-1}\Big\{\mathcal{H}'_{\mathrm{AAA}i}\bigstar\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{G}_{\mathrm{A}i}\mathcal{G}_{\mathrm{A}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\boldsymbol{U}_i^T\\
&-\phi\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\boldsymbol{U}_i^T\Big]\boldsymbol{\Sigma}^{-1}\Big)+O_p(n^{-3/2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}.
\end{aligned}\tag{S.23}$$

S.10 Score Outer Product Conditional Moments Approximation

The ith term of the Fisher information matrix of $\big(\boldsymbol{\beta},\mathrm{vech}(\boldsymbol{\Sigma})\big)$ is a 3×3 block partitioned matrix with the blocks corresponding to the various moments of pairwise outer products, conditional on $\boldsymbol{X}_i$. The relevant approximations involve repeated use of (S.14) and and keeping track of orders of magnitude.

S.10.1 Approximation of $E(\boldsymbol{S}_{\mathbf{A}i}^{\otimes2}|\boldsymbol{X}_i)$

Using (S.18), (S.14) and standard algebraic steps we have

$$\begin{aligned}
E(\boldsymbol{S}_{Ai}^{\otimes2}|\boldsymbol{X}_i)&=\boldsymbol{\Sigma}^{-1}\\
&+\phi\boldsymbol{\Sigma}^{-1}E\Big(\mathcal{H}_{\mathrm{AA}i}^{-1}-\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}-\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T\;|\boldsymbol{X}_i\Big)\boldsymbol{\Sigma}^{-1}\\
&-\phi\boldsymbol{\Sigma}^{-1}E\Big\{\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\boldsymbol{U}_i^T+\boldsymbol{U}_i\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big|\boldsymbol{X}_i\Big\}\boldsymbol{\Sigma}^{-1}\\
&+O_p(n^{-2})\mathbf{1}_{d_{\mathrm{R}}}^{\otimes2}.
\end{aligned}\tag{S.24}$$

S.10.2 Approximation of $E(\boldsymbol{S}_{\mathbf{B}i}^{\otimes 2}|\boldsymbol{X}_i)$

From (S.21) and (S.14) we obtain

$$E(\boldsymbol{S}_{Bi}^{\otimes 2}|\boldsymbol{X}_i) = \frac{1}{\phi}E\Big(\mathcal{H}_{\mathrm{BB}i} - \mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}|\boldsymbol{X}_i\Big) + O_p(1)\mathbf{1}_{d_{\mathrm{B}}}^{\otimes 2}. \tag{S.25}$$

S.10.3 Approximation of $E(\boldsymbol{S}_{\mathbf{C}i}^{\otimes 2}|\boldsymbol{X}_i)$

After some long-winded, but relatively straightforward, matrix algebra that involves application of (S.14) we have from (S.23) that

$$\begin{aligned} E(\boldsymbol{S}_{\mathrm{C}i}^{\otimes 2}|\boldsymbol{X}_i) &= \frac{1}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}} + \tfrac{\phi}{2}\boldsymbol{D}_{d_{\mathrm{R}}}^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})E\Big[2(\boldsymbol{U}_i\boldsymbol{U}_i^T)\otimes\mathcal{H}_{\mathrm{AA}i}^{-1} \\ &+\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma})\mathrm{vec}\left(\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\Big\{\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\boldsymbol{U}_i^T\Big\}\right)^T \\ &+\mathrm{vec}\left(\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\Big\{\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\boldsymbol{U}_i^T\Big\}\right)\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma})^T\Big|\boldsymbol{X}_i\Big] \\ &\times(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}} + O_p(n^{-2})\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}^{\otimes 2}. \end{aligned} \tag{S.26}$$

S.10.4 Approximation of $E(\boldsymbol{S}_{\mathbf{A}i}\boldsymbol{S}_{\mathbf{B}i}^T|\boldsymbol{X}_i)$

Multiplication of (S.18) by the transpose of (S.21), taking expectations conditional on $\boldsymbol{X}_i$ and use of (S.14) leads to

$$\begin{aligned} E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{B}i}^T|\boldsymbol{X}_i) = \boldsymbol{\Sigma}^{-1}E\Big\{&\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i} - \boldsymbol{U}_i\Big(\mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)^T \\ &+\boldsymbol{U}_i\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)^T\mathcal{H}_{\mathrm{AA}i}^{-1}\mathcal{H}_{\mathrm{AB}i}\Big|\boldsymbol{X}_i\Big\} + O_p(n^{-1})\mathbf{1}_{d_{\mathrm{R}}}\mathbf{1}_{d_{\mathrm{B}}}^T. \end{aligned} \tag{S.27}$$

S.10.5 Approximation of $E(\boldsymbol{S}_{\mathbf{A}i}\boldsymbol{S}_{\mathbf{C}i}^T|\boldsymbol{X}_i)$

An important aspect of the $E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i)$ approximation is that, even though

$$\boldsymbol{S}_{\mathrm{A}i} = O_p(1)\mathbf{1}_{d_{\mathrm{R}}} \quad \text{and} \quad \boldsymbol{S}_{\mathrm{C}i} = O_p(1)\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}$$

we can establish that

$$E(\boldsymbol{S}_{\mathrm{A}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i) = O_p(n^{-1})\mathbf{1}_{d_{\mathrm{R}}}\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}^T, \tag{S.28}$$

which indicates a degree of asymptotic orthogonality between $\boldsymbol{\beta}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}$. An illustrative cancellation, involving the leading terms of each score, is

$$E\{\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}|\boldsymbol{X}_i\}\boldsymbol{D}_{d_{\mathrm{R}}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{\Sigma}^{-1}\{E(\boldsymbol{U}_i\boldsymbol{U}_i^T|\boldsymbol{X}_i)-\boldsymbol{\Sigma}\}\boldsymbol{\Sigma}^{-1}\boldsymbol{D}_{d_{\mathrm{R}}} = \boldsymbol{O}.$$

As will be shown in Section S.12, approximation (S.28) is sufficient for (8).

S.10.6 Approximation of $E(\boldsymbol{S}_{\mathbf{B}i}\boldsymbol{S}_{\mathbf{C}i}^T|\boldsymbol{X}_i)$

Multiplication of (S.21) by the transpose of (S.23), and similar arguments, leads to

$$\begin{aligned} E\big(\boldsymbol{S}_{\mathrm{B}i}\boldsymbol{S}_{\mathrm{C}i}^T|\boldsymbol{X}_i\big) = \frac{1}{2}E\Big[\Big\{&\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i - \mathcal{H}'_{\mathrm{AAB}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1} \\ &+\mathcal{H}_{\mathrm{AB}i}^T\mathcal{H}_{\mathrm{AA}i}^{-1}\Big(\mathcal{H}'_{\mathrm{AAA}i}\bigstar\mathcal{H}_{\mathrm{AA}i}^{-1}\Big)\Big\}\mathrm{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma})^T\Big|\boldsymbol{X}_i\Big](\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\mathrm{R}}} \\ &+O_p(n^{-1})\mathbf{1}_{d_{\mathrm{B}}}\mathbf{1}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}^T. \end{aligned} \tag{S.29}$$

S.11 The Fisher Information Matrix

The Fisher information matrix of $(\boldsymbol{\beta}, \text{vech}(\boldsymbol{\Sigma}))$ is

$$I\big(\boldsymbol{\beta}, \text{vech}(\boldsymbol{\Sigma})\big) = \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{12}^T & \boldsymbol{M}_{22} \end{bmatrix} \quad \text{where} \quad \boldsymbol{M}_{11} \equiv \sum_{i=1}^{m} \begin{bmatrix} E(\boldsymbol{S}_{\text{A}i}^{\otimes 2}|\boldsymbol{X}_i) & E(\boldsymbol{S}_{\text{A}i}\boldsymbol{S}_{\text{B}i}^T|\boldsymbol{X}_i) \\ E(\boldsymbol{S}_{\text{B}i}\boldsymbol{S}_{\text{A}i}^T|\boldsymbol{X}_i) & E(\boldsymbol{S}_{\text{B}i}^{\otimes 2}|\boldsymbol{X}_i) \end{bmatrix},$$

$$\boldsymbol{M}_{12} \equiv \sum_{i=1}^{m} \begin{bmatrix} E(\boldsymbol{S}_{\text{A}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i) \\ E(\boldsymbol{S}_{\text{B}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i) \end{bmatrix} \quad \text{and} \quad \boldsymbol{M}_{22} \equiv \sum_{i=1}^{m} E(\boldsymbol{S}_{\text{C}i}^{\otimes 2}|\boldsymbol{X}_i).$$

The results of the previous section lead to high-order asymptotic approximation of the matrix $I\big(\boldsymbol{\beta}, \text{vech}(\boldsymbol{\Sigma})\big)$. In the next section we show that inversion of this approximate Fisher information matrix leads to two-term covariance matrix approximations for the maximum likelihood estimators.

S.12 Approximation of Covariance Matrices of Estimators

The dominant terms in the approximation of

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) \quad \text{and} \quad \text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big)$$

correspond to the $d_{\text{F}} \times d_{\text{F}}$ and $\frac{1}{2}d_{\text{R}}(d_{\text{R}}+1) \times \frac{1}{2}d_{\text{R}}(d_{\text{R}}+1)$ diagonal blocks of

$$I\big(\boldsymbol{\beta}, \text{vech}(\boldsymbol{\Sigma})\big)^{-1}.$$

We now treat each of these in turn in the upcoming subsections, which make extensive use of block matrix inversions. If a matrix is partitioned into four blocks $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$, then

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & (\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}$$

or, equivalently,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{bmatrix}.$$

Another result that is repeatedly used in the following subsections is

$$(\boldsymbol{A}-\boldsymbol{B})^{-1} = \sum_{k=0}^{\infty}(\boldsymbol{A}^{-1}\boldsymbol{B})^k\boldsymbol{A}^{-1}$$

for $\boldsymbol{A}$ and $\boldsymbol{B}$ invertible matrices of the same size and such that the spectral radius of $\boldsymbol{A}^{-1}\boldsymbol{B}$ is less than 1.

S.12.1 Two-Term Approximation of $\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big)$

The dominant terms of $\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big)$ correspond to

$$\text{the upper left } d_{\text{F}} \times d_{\text{F}} \text{ block of } I\big(\boldsymbol{\beta}, \text{vech}(\boldsymbol{\Sigma})\big)^{-1} = \big(\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T\big)^{-1}.$$

Based on (S.24), (S.25) and (S.27) we have

$$\boldsymbol{M}_{11} = \begin{bmatrix} m\boldsymbol{\Sigma}^{-1} - \dfrac{\phi m}{n}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{K}}_{\text{AA}}\boldsymbol{\Sigma}^{-1} + O_p(mn^{-2})\boldsymbol{1}_{d_{\text{R}}}^{\otimes 2} & m\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{K}}_{\text{AB}} + O_p(mn^{-1})\boldsymbol{1}_{d_{\text{R}}}\boldsymbol{1}_{d_{\text{B}}}^T \\ m\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{K}}_{\text{AB}}^T + O_p(mn^{-1})\boldsymbol{1}_{d_{\text{B}}}\boldsymbol{1}_{d_{\text{R}}}^T & \dfrac{mn}{\phi}\boldsymbol{\mathcal{K}}_{\text{BB}} + O_p(m)\boldsymbol{1}_{d_{\text{B}}}^{\otimes 2} \end{bmatrix}$$

where

$$\mathcal{K}_{\text{AA}} \equiv \frac{n}{m}\sum_{i=1}^{m} E\Big\{\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\text{AA}i}^{-1} + \mathcal{H}_{\text{AA}i}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T - \mathcal{H}_{\text{AA}i}^{-1}$$
$$+\mathcal{H}_{\text{AA}i}^{-1}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)\boldsymbol{U}_i^T + \boldsymbol{U}_i\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T\mathcal{H}_{\text{AA}i}^{-1}\Big|\boldsymbol{X}_i\Big\},$$

$$\mathcal{K}_{\text{AB}} \equiv \frac{1}{m}\sum_{i=1}^{m} E\Big\{\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i} - \boldsymbol{U}_i\Big(\mathcal{H}'_{\text{AAB}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T$$
$$+\boldsymbol{U}_i\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}_{\text{AA}i}^{-1}\Big)^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\Big|\boldsymbol{X}_i\Big\}$$

and $$\mathcal{K}_{\text{BB}} \equiv \frac{1}{mn}\sum_{i=1}^{m} E\Big(\mathcal{H}_{\text{BB}i} - \mathcal{H}_{\text{AB}i}^T\mathcal{H}_{\text{AA}i}^{-1}\mathcal{H}_{\text{AB}i}\Big|\boldsymbol{X}_i\Big)$$

are matrices with all entries being $O_p(1)$. As consequences of (S.26), (S.28) and (S.29) we have

$$\boldsymbol{M}_{22}^{-1} = O_p(m^{-1})\mathbf{1}_{d_{\text{R}}(d_{\text{R}}+1)/2}^{\otimes 2} \quad \text{and} \quad \boldsymbol{M}_{12} = \begin{bmatrix} O_p(mn^{-1})\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{R}}(d_{\text{R}}+1)/2}^T \\ O_p(m)\mathbf{1}_{d_{\text{B}}}\mathbf{1}_{d_{\text{R}}(d_{\text{R}}+1)/2}^T \end{bmatrix}. \tag{S.30}$$

Therefore

$$\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T = \begin{bmatrix} O_p(mn^{-2})\mathbf{1}_{d_{\text{R}}}^{\otimes 2} & O_p(mn^{-1})\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{B}}}^T \\ O_p(mn^{-1})\mathbf{1}_{d_{\text{B}}}\mathbf{1}_{d_{\text{R}}}^T & O_p(m)\mathbf{1}_{d_{\text{B}}}^{\otimes 2} \end{bmatrix}.$$

From these results for $\boldsymbol{M}_{11}$ and $\boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T$, it follows that

$$\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T =$$
$$\begin{bmatrix} m\boldsymbol{\Sigma}^{-1} - \frac{\phi m}{n}\boldsymbol{\Sigma}^{-1}\mathcal{K}_{\text{AA}}\boldsymbol{\Sigma}^{-1} + O_p(mn^{-2})\mathbf{1}_{d_{\text{R}}}^{\otimes 2} & m\boldsymbol{\Sigma}^{-1}\mathcal{K}_{\text{AB}} + O_p(mn^{-1})\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{B}}}^T \\ m\boldsymbol{\Sigma}^{-1}\mathcal{K}_{\text{AB}}^T + O_p(mn^{-1})\mathbf{1}_{d_{\text{B}}}\mathbf{1}_{d_{\text{R}}}^T & \frac{mn}{\phi}\mathcal{K}_{\text{BB}} + O_p(m)\mathbf{1}_{d_{\text{B}}}^{\otimes 2} \end{bmatrix}.$$

The upper left $d_{\text{R}} \times d_{\text{R}}$ block of $\big(\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T\big)^{-1}$ is

$$\Big\{m\boldsymbol{\Sigma}^{-1} - \frac{\phi m}{n}\boldsymbol{\Sigma}^{-1}\big(\mathcal{K}_{\text{AA}} + \mathcal{K}_{\text{AB}}\mathcal{K}_{\text{BB}}^{-1}\mathcal{K}_{\text{AB}}^T\big)\boldsymbol{\Sigma}^{-1} + O_p(mn^{-2})\mathbf{1}_{d_{\text{R}}}^{\otimes 2}\Big\}^{-1}$$
$$= \frac{1}{m}\Big\{\boldsymbol{I} - \frac{\phi}{n}\big(\mathcal{K}_{\text{AA}} + \mathcal{K}_{\text{AB}}\mathcal{K}_{\text{BB}}^{-1}\mathcal{K}_{\text{AB}}^T\big)\boldsymbol{\Sigma}^{-1} + O_p(n^{-2})\mathbf{1}_{d_{\text{R}}}^{\otimes 2}\Big\}^{-1}\boldsymbol{\Sigma}$$
$$= \frac{\boldsymbol{\Sigma}}{m} + \frac{\phi\big(\mathcal{K}_{\text{AA}} + \mathcal{K}_{\text{AB}}\mathcal{K}_{\text{BB}}^{-1}\mathcal{K}_{\text{AB}}^T\big)}{mn} + O_p(m^{-1}n^{-2})\mathbf{1}_{d_{\text{R}}}^{\otimes 2}.$$

The upper right $d_{\text{R}} \times d_{\text{B}}$ block of $\big(\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T\big)^{-1}$ is

$$-\left\{\frac{\boldsymbol{\Sigma}}{m} + O_p(m^{-1}n^{-1})\mathbf{1}_{d_{\text{R}}}^{\otimes 2}\right\}\left\{m\boldsymbol{\Sigma}^{-1}\mathcal{K}_{\text{AB}} + O_p(mn^{-1})\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{B}}}^T\right\}\left\{\frac{mn}{\phi}\mathcal{K}_{\text{BB}} + O_p(m)\mathbf{1}_{d_{\text{B}}}^{\otimes 2}\right\}^{-1}$$
$$= -\frac{\phi\mathcal{K}_{\text{AB}}\mathcal{K}_{\text{BB}}^{-1}}{mn} + O_p(m^{-1}n^{-2})\mathbf{1}_{d_{\text{R}}}\mathbf{1}_{d_{\text{B}}}^T.$$

The lower right $d_{\text{B}} \times d_{\text{B}}$ block of $\big(\boldsymbol{M}_{11} - \boldsymbol{M}_{12}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{12}^T\big)^{-1}$ is

$$\frac{\phi\mathcal{K}_{\text{BB}}^{-1}}{mn} + O_p(m^{-1}n^{-2})\mathbf{1}_{d_{\text{B}}}^{\otimes 2}.$$

Therefore,

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\begin{bmatrix}\boldsymbol{\Sigma}^0 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix} + \frac{\phi}{mn}\begin{bmatrix}(\boldsymbol{\mathcal{K}}^0_{\text{AA}})^{-1} & (\boldsymbol{\mathcal{K}}^0_{\text{AA}})^{-1}\boldsymbol{\mathcal{K}}^0_{\text{AB}}\\ (\boldsymbol{\mathcal{K}}^0_{\text{AB}})^T(\boldsymbol{\mathcal{K}}^0_{\text{AA}})^{-1} & \boldsymbol{\mathcal{K}}^0_{\text{BB}} + (\boldsymbol{\mathcal{K}}^0_{\text{AB}})^T(\boldsymbol{\mathcal{K}}^0_{\text{AA}})^{-1}\boldsymbol{\mathcal{K}}^0_{\text{AB}}\end{bmatrix}^{-1}$$
$$+O_p(m^{-1}n^{-2})\mathbf{1}^{\otimes 2}_{d_{\text{F}}}$$

where, for example, $\boldsymbol{\mathcal{K}}^0_{\text{AA}}$ is the $\boldsymbol{\mathcal{K}}_{\text{AA}}$ quantity with $\boldsymbol{\beta}$ set to $\boldsymbol{\beta}^0$ and $\boldsymbol{\Sigma}$ set to $\boldsymbol{\Sigma}^0$.

S.12.2 Two-Term Approximation of $\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big)$

The dominant terms of $\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big)$ correspond to

$$\text{the lower right } \frac{1}{2}d_{\text{R}}(d_{\text{R}}+1)\times\frac{1}{2}d_{\text{R}}(d_{\text{R}}+1)\text{ block of } I\big(\boldsymbol{\beta},\text{vech}(\boldsymbol{\Sigma})\big)^{-1}$$
$$= \big(\boldsymbol{M}_{22} - \boldsymbol{M}_{12}^T\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}\big)^{-1}.$$

From (S.26)

$$\sum_{i=1}^m E(\boldsymbol{S}_{\text{C}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i) = \frac{m}{2}\boldsymbol{D}^T_{d_{\text{R}}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{R}}} - \frac{m\phi}{n}\boldsymbol{D}^T_{d_{\text{R}}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{\mathcal{K}}_{\text{CC}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{R}}}$$
$$+O_p(mn^{-2})\mathbf{1}^{\otimes 2}_{d_{\text{R}}(d_{\text{R}}+1)/2}$$

with the following $O_p(1)$ matrix:

$$\boldsymbol{\mathcal{K}}_{\text{CC}} = \frac{n}{m}\sum_{i=1}^m E\Big[\frac{1}{2}\text{vec}(\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T)\text{vec}\Big(\mathcal{H}^{-1}_{\text{AA}i}\boldsymbol{\Sigma}^{-1}\Big\{\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}^{-1}_{\text{AA}i}\Big)\boldsymbol{U}_i^T\Big\}\Big)^T$$
$$+\frac{1}{2}\text{vec}\Big(\mathcal{H}^{-1}_{\text{AA}i}\boldsymbol{\Sigma}^{-1}\Big\{\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}^{-1}_{\text{AA}i}\Big)\boldsymbol{U}_i^T\Big\}\Big)\text{vec}(\boldsymbol{\Sigma}-\boldsymbol{U}_i\boldsymbol{U}_i^T)^T$$
$$-(\boldsymbol{U}_i\boldsymbol{U}_i^T)\otimes(\mathcal{H}^{-1}_{\text{AA}i})\Big|\boldsymbol{X}_i\Big].$$

Next, note that

$$\boldsymbol{M}_{11}^{-1} = \begin{bmatrix}O_p(m^{-1})\mathbf{1}^{\otimes 2}_{d_{\text{R}}} & O_p\{(mn)^{-1}\}\mathbf{1}_{d_{\text{R}}}\mathbf{1}^T_{d_{\text{B}}}\\ O_p\{(mn)^{-1}\}\mathbf{1}_{d_{\text{B}}}\mathbf{1}^T_{d_{\text{R}}} & O_p\{(mn)^{-1}\}\mathbf{1}^{\otimes 2}_{d_{\text{B}}}\end{bmatrix}. \tag{S.31}$$

Given the orders of magnitude in (S.30) and (S.31), from expansion of $\boldsymbol{M}_{12}^T\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}$ it is apparent that its dominant $O_p(m/n)$ contribution is from

$$\Bigg\{\sum_{i=1}^m E(\boldsymbol{S}_{\text{B}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i)\Bigg\}^T\Bigg\{\sum_{i=1}^m E(\boldsymbol{S}_{\text{B}i}\boldsymbol{S}_{\text{B}i}^T|\boldsymbol{X}_i)\Bigg\}^{-1}\sum_{i=1}^m E(\boldsymbol{S}_{\text{B}i}\boldsymbol{S}_{\text{C}i}^T|\boldsymbol{X}_i)$$
$$= \frac{\phi m}{n}\boldsymbol{D}^T_{d_{\text{R}}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{\mathcal{K}}_{\text{BC}}\boldsymbol{\mathcal{K}}_{\text{BB}}^{-1}\boldsymbol{\mathcal{K}}_{\text{BC}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{D}_{d_{\text{R}}} + o_p(m/n)\mathbf{1}^{\otimes 2}_{d_{\text{R}}(d_{\text{R}}+1)/2}$$

where

$$\boldsymbol{\mathcal{K}}_{\text{BC}} \equiv \frac{1}{2m}\sum_{i=1}^m E\Big[\Big\{\mathcal{H}^T_{\text{AB}i}\mathcal{H}^{-1}_{\text{AA}i}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i - \mathcal{H}'_{\text{AAB}i}\bigstar\mathcal{H}^{-1}_{\text{AA}i}$$
$$+\mathcal{H}^T_{\text{AB}i}\mathcal{H}^{-1}_{\text{AA}i}\Big(\mathcal{H}'_{\text{AAA}i}\bigstar\mathcal{H}^{-1}_{\text{AA}i}\Big)\Big\}\Big|\boldsymbol{X}_i\Big]\text{vec}(\boldsymbol{U}_i\boldsymbol{U}_i^T-\boldsymbol{\Sigma})^T$$

is a matrix with all entries being $O_p(1)$. Hence, if we let $\boldsymbol{\mathcal{A}} \equiv \frac{1}{2}\boldsymbol{D}^T_{d_{\text{R}}}\big((\boldsymbol{\Sigma}^0)^{-1}\otimes(\boldsymbol{\Sigma}^0)^{-1}\big)\boldsymbol{D}_{d_{\text{R}}}$ and

$$\boldsymbol{\mathcal{B}} \equiv \boldsymbol{D}^T_{d_{\text{R}}}\big((\boldsymbol{\Sigma}^0)^{-1}\otimes(\boldsymbol{\Sigma}^0)^{-1}\big)\big\{\boldsymbol{\mathcal{K}}^0_{\text{CC}} + (\boldsymbol{\mathcal{K}}^0_{\text{BC}})^T(\boldsymbol{\mathcal{K}}^0_{\text{BB}})^{-1}\boldsymbol{\mathcal{K}}^0_{\text{BC}}\big\}\big((\boldsymbol{\Sigma}^0)^{-1}\otimes(\boldsymbol{\Sigma}^0)^{-1}\big)\boldsymbol{D}_{d_{\text{R}}}$$

then

$$\mathrm{Cov}\big(\mathrm{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\left(\boldsymbol{\mathcal{A}} - \frac{\phi}{n}\boldsymbol{\mathcal{B}}\right)^{-1} + o_p\{(mn)^{-1}\}\mathbf{1}^{\otimes 2}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}$$

$$= \frac{1}{m}\boldsymbol{\mathcal{A}}^{-1} + \frac{\phi}{mn}\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{A}}^{-1} + o_p\{(mn)^{-1}\}\mathbf{1}^{\otimes 2}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}.$$

From (S.5),

$$\boldsymbol{\mathcal{A}}^{-1} = 2\boldsymbol{D}^{+}_{d_{\mathrm{R}}}(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}^{+T}_{d_{\mathrm{R}}}.$$

To simplify $\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{A}}^{-1}$, we use (S.4) and (S.7) to obtain

$$\mathrm{Cov}\big(\mathrm{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{2\boldsymbol{D}^{+}_{d_{\mathrm{R}}}(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}^{+T}_{d_{\mathrm{R}}}}{m} + \frac{4\phi\boldsymbol{D}^{+}_{d_{\mathrm{R}}}\{\boldsymbol{\mathcal{K}}^0_{\mathrm{CC}} + (\boldsymbol{\mathcal{K}}^0_{\mathrm{BC}})^T(\boldsymbol{\mathcal{K}}^0_{\mathrm{BB}})^{-1}\boldsymbol{\mathcal{K}}^0_{\mathrm{BC}}\}\boldsymbol{D}^{+T}_{d_{\mathrm{R}}}}{mn} + O_p(m^{-1}n^{-2})\mathbf{1}^{\otimes 2}_{d_{\mathrm{R}}(d_{\mathrm{R}}+1)/2}. \tag{S.32}$$

S.13 Population Forms of Covariance Matrix Second Terms

In the previous section, the second terms of the asymptotic covariance matrices of $\widehat{\boldsymbol{\beta}}$ and $\mathrm{vech}(\widehat{\boldsymbol{\Sigma}})$ are stochastic. However, under relatively mild moment conditions such as assumption (A3) of Jiang *et al.* (2022), these terms converge in probability to deterministic population forms. In this section we determine these limiting forms.

A re-writing of the $\boldsymbol{\mathcal{K}}_{\mathrm{AA}}$ quantity is

$$\begin{aligned}\boldsymbol{\mathcal{K}}_{\mathrm{AA}} \equiv \frac{1}{m}\sum_{i=1}^{m} E\Big[&\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1} + \big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T - \big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\\ &+\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\Big\{\big(\tfrac{1}{n}\mathcal{H}'_{\mathrm{AAA}i}\big)\bigstar\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\big)\Big\}\boldsymbol{U}_i^T\\ &+\boldsymbol{U}_i\Big\{\big(\tfrac{1}{n}\mathcal{H}'_{\mathrm{AAA}i}\big)\bigstar\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\Big\}^T\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}\big)^{-1}\Big|\boldsymbol{X}_i\Big].\end{aligned}$$

Since

$$E\big(\tfrac{1}{n}\mathcal{H}_{\mathrm{AA}i}|\boldsymbol{X}_i\big) \xrightarrow{P} \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i) \quad \text{and} \quad E\big(\tfrac{1}{n}\mathcal{H}'_{\mathrm{AAA}i}|\boldsymbol{X}_i\big) \xrightarrow{P} \boldsymbol{\Omega}'_{\mathrm{AAA}}(\boldsymbol{U}_i)$$

we have, under relatively mild conditions (see e.g. Lemma A1 of Jiang *et al.*, 2022),

$$\begin{aligned}\boldsymbol{\mathcal{K}}_{\mathrm{AA}} \xrightarrow{P} \frac{1}{m}\sum_{i=1}^{m} E\Big[&\boldsymbol{U}_i\boldsymbol{U}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1} + \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^T - \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\\ &+\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\Big\{\boldsymbol{\Omega}'_{\mathrm{AAA}}(\boldsymbol{U}_i)\bigstar\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\Big\}\boldsymbol{U}_i^T\\ &+\boldsymbol{U}_i\Big\{\boldsymbol{\Omega}'_{\mathrm{AAA}}(\boldsymbol{U}_i)\bigstar\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\Big\}^T\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U}_i)^{-1}\Big]\\ = E\Big[&\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1} + \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\boldsymbol{U}^T - \boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\\ &+\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\Big\{\boldsymbol{\Omega}'_{\mathrm{AAA}}(\boldsymbol{U})\bigstar\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\Big\}\boldsymbol{U}^T + \boldsymbol{U}\Big\{\boldsymbol{\Omega}'_{\mathrm{AAA}}(\boldsymbol{U})\bigstar\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\Big\}^T\boldsymbol{\Omega}_{\mathrm{AA}}(\boldsymbol{U})^{-1}\Big]\\ = \boldsymbol{\Lambda}_{\mathrm{AA}}&\end{aligned}$$

where $\boldsymbol{\Lambda}_{\mathrm{AA}}$ is as defined in Section 3.1. Analogous arguments lead to

$$\boldsymbol{\mathcal{K}}_{\mathrm{AB}} \xrightarrow{P} \boldsymbol{\Lambda}_{\mathrm{AB}}, \quad \boldsymbol{\mathcal{K}}_{\mathrm{BB}} \xrightarrow{P} E\{\boldsymbol{\Psi}_6(\boldsymbol{U})\}, \quad \boldsymbol{\mathcal{K}}_{\mathrm{BC}} \xrightarrow{P} \boldsymbol{\Phi} \ \text{ and } \ \boldsymbol{D}^{+}_{d_{\mathrm{R}}}\boldsymbol{\mathcal{K}}_{\mathrm{CC}}\boldsymbol{D}^{+T}_{d_{\mathrm{R}}} \xrightarrow{P} \frac{1}{2}E\big\{\boldsymbol{\Psi}_9(\boldsymbol{U}) - 2\boldsymbol{\Psi}_8(\boldsymbol{U})\big\}$$

where $\boldsymbol{\Psi}_8(\boldsymbol{U})$, $\boldsymbol{\Psi}_9(\boldsymbol{U})$ and $\boldsymbol{\Lambda}_{\mathrm{AB}}$ are as defined in Section 3.1. It follows that the deterministic forms of the order $(mn)^{-1}$ terms match those stated in (8).

S.14 The Gaussian Response Special Case

For the Gaussian response special case of (3) the two-term covariance matrix expressions simplify considerably. The main reason is that, for the Gaussian case, $b''(x) = 1$ and $b'''(x) = 0$. These facts imply that

$$\boldsymbol{\Omega}_{\text{AA}}(\boldsymbol{U}) = E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T), \quad \boldsymbol{\Omega}_{\text{AB}}(\boldsymbol{U}) = E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T), \quad \boldsymbol{\Omega}_{\text{BB}}(\boldsymbol{U}) = E(\boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{B}}^T)$$

and all entries of the three-dimensional arrays $\boldsymbol{\Omega}'_{\text{AAA}}(\boldsymbol{U})$ and $\boldsymbol{\Omega}'_{\text{AAB}}(\boldsymbol{U})$ are exactly zero.

S.14.1 The $\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big)$ Approximation

For the Gaussian response situation

$$\boldsymbol{\Lambda}_{\text{AA}} = E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)^{-1}, \quad \boldsymbol{\Lambda}_{\text{AB}} = E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)^{-1}E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T)$$

and

$$E\{\boldsymbol{\Psi}_6(\boldsymbol{U})\} = E(\boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{B}}^T) - E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T)^T E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)^{-1}E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T).$$

Therefore,

$$\begin{bmatrix} \boldsymbol{\Lambda}_{\text{AA}}^{-1} & \boldsymbol{\Lambda}_{\text{AA}}^{-1}\boldsymbol{\Lambda}_{\text{AB}} \\ \boldsymbol{\Lambda}_{\text{AB}}^T\boldsymbol{\Lambda}_{\text{AA}}^{-1} & \boldsymbol{\Lambda}_{\text{AB}}^T\boldsymbol{\Lambda}_{\text{AA}}^{-1}\boldsymbol{\Lambda}_{\text{AB}} + E\big\{\boldsymbol{\Psi}_6(\boldsymbol{U})\big\} \end{bmatrix} = E\begin{bmatrix} \boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T & \boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{B}}^T \\ \boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{A}}^T & \boldsymbol{X}_{\text{B}}\boldsymbol{X}_{\text{B}}^T \end{bmatrix} = E(\boldsymbol{X}\boldsymbol{X}^T).$$

Hence, for the Gaussian special case

$$\text{Cov}\big(\widehat{\boldsymbol{\beta}}|\boldsymbol{\mathcal{X}}\big) = \frac{1}{m}\begin{bmatrix} \boldsymbol{\Sigma}^0 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} + \frac{\phi\big\{E(\boldsymbol{X}\boldsymbol{X}^T)\big\}^{-1}\{1+o_p(1)\}}{mn}.$$

This result generalises the two-term expansion of $\text{Var}\big(\widehat{\beta}_{\text{A}}|\mathcal{X}\big)$ provided in Section 3.5 of McCulloch *et al.* (2008) for the $d_{\text{R}} = d_{\text{B}} = 1$ and $\boldsymbol{X}_{\text{A}} = 1$ special case.

S.14.2 The $\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big)$ Approximation

As shown in, for example, Section 4.3 of Wand (2002) there is exact orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in the Gaussian case. This means that $\boldsymbol{\Phi} = \boldsymbol{O}$ and, hence, the second term of $\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big)$ is

$$\frac{2\phi}{mn}E\Big\{\text{vech}(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T)\psi_4(\boldsymbol{U})^T + \psi_4(\boldsymbol{U})\text{vech}(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T)^T - 2\boldsymbol{\Psi}_8(\boldsymbol{U})\Big\} \tag{S.33}$$

where $\psi_4(\boldsymbol{U})$ and $\boldsymbol{\Psi}_8(\boldsymbol{U})$ simplify to

$$\psi_4(\boldsymbol{U}) = \boldsymbol{D}_{d_{\text{R}}}^{+}\text{vec}\Big(\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T\big)\Big)$$

and

$$\boldsymbol{\Psi}_8(\boldsymbol{U}) = \boldsymbol{D}_{d_{\text{R}}}^{+}\big[(\boldsymbol{U}\boldsymbol{U}^T)\otimes\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\big]\boldsymbol{D}_{d_{\text{R}}}^{+T}.$$

We immediately have

$$E\{\boldsymbol{\Psi}_8(\boldsymbol{U})\} = \boldsymbol{D}_{d_{\text{R}}}^{+}\big[\boldsymbol{\Sigma}\otimes\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\big]\boldsymbol{D}_{d_{\text{R}}}^{+T}.$$

The reduction of the other expectations in (S.33) is less immediate and benefits from Theorem 4.3(iv) of Magnus & Neudecker (1979) as well as (S.2). However, such a pathway leads to

$$E\Big\{\text{vech}(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T)\psi_4(\boldsymbol{U})^T + \psi_4(\boldsymbol{U})\text{vech}(\boldsymbol{\Sigma}-\boldsymbol{U}\boldsymbol{U}^T)^T\Big\} = 4\boldsymbol{D}_{d_{\text{R}}}^{+}\big[\boldsymbol{\Sigma}\otimes\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\big]\boldsymbol{D}_{d_{\text{R}}}^{+T}.$$

On combining the components of (S.33) we arrive at

$$\text{Cov}\big(\text{vech}(\widehat{\boldsymbol{\Sigma}})|\boldsymbol{\mathcal{X}}\big) = \frac{2\boldsymbol{D}_{d_{\text{R}}}^{+}(\boldsymbol{\Sigma}^0\otimes\boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\text{R}}}^{+T}}{m} + \frac{4\phi\boldsymbol{D}_{d_{\text{R}}}^{+}\big[\boldsymbol{\Sigma}^0\otimes\{E(\boldsymbol{X}_{\text{A}}\boldsymbol{X}_{\text{A}}^T)\}^{-1}\big]\boldsymbol{D}_{d_{\text{R}}}^{+T}\{1+o_p(1)\}}{mn}.$$

S.15 Additional Simulation Exercise Figure

Figure S.1 refers to the simulation exercise described in Section 5 and compares the empirical coverages of confidence intervals with advertised levels of 95% for the parameters of (8) that are not affected by second term improvement. It is clear from Figure S.1 that the simple one-term asymptotic variances lead to good coverages for β_2^0, β_3^0 and β_4^0, even for lower sample size situations.

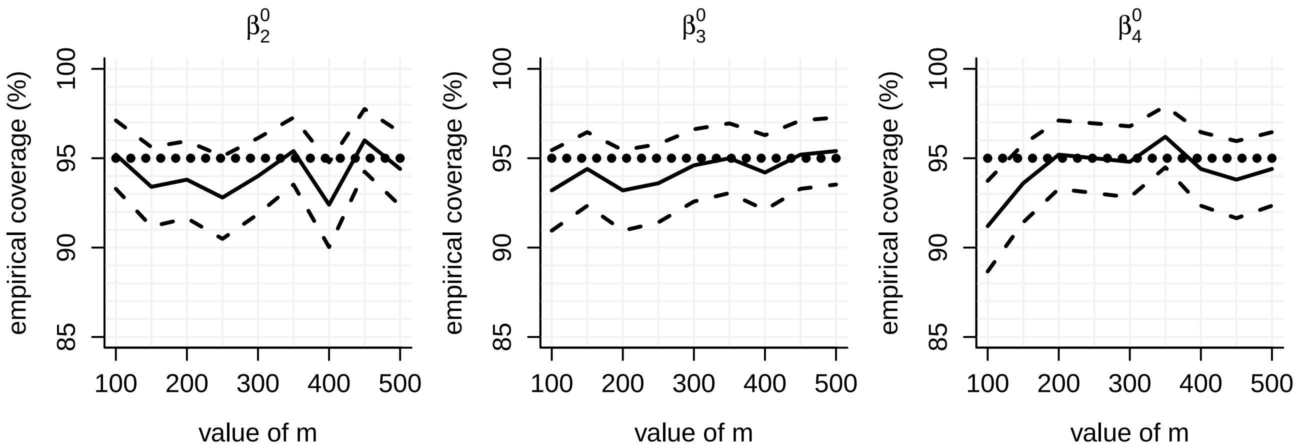

Figure S.1: *Empirical coverage of confidence intervals from the simulation exercise described in Section 5. Each panel corresponds to a fixed effect model parameter that is not impacted by second term asymptotic improvements. The advertised coverage level is fixed at 95% and is indicated by a horizontal dotted line in each panel. The solid curves show, dependent on the number of groups m, the empirical coverage levels for confidence intervals that use a one-term asymptotic variance approximation. The dashed curves correspond to plus and minus two standard errors of the sample proportions. The within-group sample size, n, is fixed at $m/10$.*

References

Jiang, J., Wand, M.P. & Bhaskaran, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *Journal of the Royal Statistical Society, Series B*, **84**, 55–82.

Magnus, J.R. and Neudecker, H. (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, **7**, 381–394.

Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus. Revised Edition.* Chichester, U.K.: John Wiley & Sons.

Miyata, Y. (2004). Fully exponential Laplace approximation using asymptotic modes. *Journal of the American Statistical Association*, **99**, 1037–1049.

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective.* Singapore: World Scientific Publishing Company.

Wand, M.P. (2002). Vector differential calculus in statistics. *The American Statistician*, **56**, 55–62.